THE UNIVERSITY *of* EDINBURGH

A thesis submitted for the degree of
Doctor of Philosophy by

Urszula McCaughan MSc (BSc)
S0680907

October 2014

Characterisation of the eukaryotic ribosome biogenesis factors, Nob1, Dim2, and Tsr1, and their interactions with RNA

Wellcome Trust Centre for Cell Biology
University of Edinburgh

DECLARATION

The work presented in this thesis was performed by the author. Any contributions made by others are clearly indicated. None of the work presented here has been submitted for another degree or professional qualification.

Signed,

Urszula McCaughan

TABLE OF CONTENTS

i. Abbreviations i

ii. Lay Summary iv

iii. Abstract v

1 Introduction 1

1.1 Ribosomal Function and Abnormalities 1
1.2 Overview of the Early rRNA Processing in Eukaryotes 2
1.3 Overview of the Late 18S rRNA Processing in Eukaryotes 3
1.4 Previous Studies Performed on Nob1 8
1.5 Previous Studies Performed on Dim2 12
1.6 Previous Studies Performed on Tsr1 16
1.7 Model Organisms 20
1.8 Aims of Project 20

2 Materials and Methods 22

1.1 Preliminary Cloning 22
1.2 Conventional Cloning 22
1.3 Ligase-Independent Cloning (LIC) 25
1.4 Gibson Assembly Cloning 27
1.5 Expression Trials 28
1.6 Fungal Protein Expression and Purification 29
1.7 Human Protein Expression and Purification 35
1.8 rRNA *In Vitro* Transcription 39
1.9 Primer Extension to Determine rRNA Secondary Structure 42
1.10 Complex Formation and Purification 44
1.11 Biochemical Assays 45
1.12 Dim2 Modelling 52
1.13 Calculating the Predicted Hydrodynamic Radii 53
1.14 Electron Microscopy (EM) 53
1.15 Crystallisation 54
1.16 Standard Laboratory Protocols 60

3 Characterisation of Nob1 65

3.1 Introduction 65
3.2 Cloning and Expression Trials 66
3.3 Protein Purification 70
3.4 The ScNob1 Protein is Unlikely to be Monomeric 83
3.5 Determination of Protein Stability by Thermal Denaturation 85
3.6 Identification of Stable Domains by Limited Proteolysis 88
3.7 Crystallisation Trials 88

3.8 Discussion 90
3.9 Concluding remarks 96
4 Investigating Nob1-RNA Interactions 97
4.1 Introduction 97
4.2 Nob1 Does Not Interact with Helix 40 RNA *in vitro* 98
4.3 Nob1 Does Interact with D-site Derived Sequences 111
4.4 Nob1 binds to an ITS1 loop sequence 130
4.5 Nob1 can Cleave the D-site *in vitro* 132
4.6 Crystallisation Trials using Protein-RNA Mixtures 135
4.7 Purification of ScNob1-D site RNA complex 135
4.8 Discussion 138
4.9 Concluding Remarks 145
5 Characterisation of Dim2 and Its Interactions 147
5.1 Introduction 147
5.2 ScDim2 Modelling Based on the Archaeal Structure 148
5.3 Cloning and Expression Trials 151
5.4 Protein Purification 156
5.5 Oligomerisation State of Dim2 Remains in Question 162
5.6 Determination of Protein Stability by Thermal Denaturation 164
5.7 Identification of Stable Domains by Limited Proteolysis 166
5.8 Testing Interactions with Different Binding Candidates 169
5.9 Dim2 Interacts with D-site Derived Sequences 177
5.10 Crystallisation Trials 189
5.11 Discussion 189
5.12 Concluding remarks 195
6 Optimisation of Tsr1 Crystals 196
6.1 Introduction 196
6.2 Native Crystal Optimisation 196
6.3 Native Data Processing 201
6.4 Phasing using Molecular Replacement and Heavy Atom Soaking 202
6.5 Purification of Seleno-Methionine labelled Tsr1 207
6.6 Optimisation of Seleno-Methionine Crystals 210
6.7 Discussion 212
6.8 Concluding remarks 216
6.9 Update 216
7 Conclusions and Future Prospects 218
7.1 Recombinant Ribosome Biogenesis Factors were Successfully Purified and Characterised 218
7.2 Insight into the 20S pre-rRNA Processing has been Obtained 220
7.3 Crystallisation of the Factors has been Successful 221
7.4 A System for Observing D site Cleavage *In Vitro* has been Optimised 222

7.5 Concluding Remarks 223
8 Acknowledgements 225
9 Appendices 226
9.1 Appendix I: Primers 226
9.2 Appendix II: LIC Vectors 229
9.3 Appendix III: Additional Plasmids 231
9.4 Appendix IV: Transcription Sequences 232
9.5 Appendix V: Tsr1 MAD Phasing 233
10 References 235

TABLE OF FIGURES

Figure 1.1: Summary of the cleavage steps involved in making the mature rRNAs in *S. cerevisiae* and *H. sapiens*. 4

Figure 1.2 (Campbell & Karbstein, 2011): Some main assembly factors involved in SSU rRNA processing. 6

Figure 1.3 (Strunk et al, 2011): Positioning of some of the assembly factors on the pre-40S ribosome as shown by EM at 18 Å resolution. 7

Figure 1.4: Sequence alignment of Nob1 from different organisms (obtained using ClustalW and Jalview). 9

Figure 1.5 (Lebaron et al, 2012): Model for D site cleavage. 11

Figure 1.6 (Veith et al, 2011): The NMR-determined structures of the two functional domains of an archaeal (from *Pyrococcus horikoshii*) Nob1 protein (PDB ID: 2LCQ). 12

Figure 1.7: Sequence alignments of Nob1 and Dim2 from yeast and archaea (obtained using ClustalW and Jalview). 13

Figure 1.8: Sequence alignment of Dim2 from different organisms (obtained using ClustalW and Jalview). 14

Figure 1.9: The general fold of archaeal Dim2 (blue) in complex with bacterial rRNA (black and red) (PDB: 3AEV). 15

Figure 1.10: Secondary structure of the 20S rRNA from *S. cerevisiae* (obtained from http://www.rna.ccbb.utexas.edu/). 17

Figure 1.11: Sequence alignment of Tsr1 from different organisms (obtained using ClustalW and Jalview). 19

Figure 2.1: Cloning of transcription sequences for RNAs with secondary structures. 24

Figure 3.1: Expression trials for His-tagged wild type (WT) ScNob1 and the D15N mutant obtained from the Tollervey lab. 67

Figure 3.2: Expression trials for the different Nob1 constructs created during this project. 68

Figure 3.3: Alignment and expression of the ScNob1 loop-out mutants created by Dr. Atlanta Cook. 69

Figure 3.4: Expression trials for CtNob1 in different cell lines. 70

Figure 3.5: Initial purification of ScNob1. 71

Figure 3.6: Optimised purification of ScNob1. 72

Figure 3.7: Initial purification of CtNob1. 74

Figure 3.8: Optimised purification of CtNob1. 75

Figure 3.9: Purification of N-terminally tagged HsNob1. 76

Figure 3.10: Purification of C-terminally tagged HsNob1..... 77
Figure 3.11: Initial purification of the ScNob1 D15N mutant. 79
Figure 3.12: SEC results for the ScNob1 D15N mutant when sonication is used after cell lysis..... 80
Figure 3.13: Purification results for the ScNob1 ΔL4 mutant. 81
Figure 3.14: SEC profile for the ScNob1ΔL3 mutant. 82
Figure 3.15: ScNob1 and ScNob1ΔL4 SEC results compared to size standards. 83
Figure 3.16: ScNob1 molar mass plot obtained from SEC-MALS analysis. 84
Figure 3.17: Thermal denaturation results for ScNob1. 86
Figure 3.18: Thermal denaturation results for CtNob1. 87
Figure 3.19: Limited proteolysis and mass spectrometry analysis of CtNob1. 89
Figure 3.20: Initial crystal hits obtained with ScNob1ΔL2 and ΔL3. 90
Figure 3.21: The two possibilities for domain organisation of Nob1. 95
Figure 4.1: Expression trial of T7 polymerase in M15 cells in Superbroth and 2xTY media. 98
Figure 4.2: Purification of the T7 polymerase using a previously established protocol. 100
Figure 4.3: An *in vitro* transcription test using the commercial MEGAshortscript kit and the purified T7 polymerase. 101
Figure 4.4: *In vitro* transcription of the different H40 RNAs cloned..... 103
Figure 4.5: Primer extension results using the three different H40 RNAs. 104
Figure 4.6: Semi-analytical SEC of ScNob1 and ScH40. 106
Figure 4.7: Semi-analytical SEC of CtNob1 and CtH40. 107
Figure 4.8: Semi-analytical SEC of CtNob1 and ScH40. 108
Figure 4.9: Semi-analytical SEC of CtNob1 and ScH40 intermediate RNA. 109
Figure 4.10: EMSA results for ScNob1 and CtNob1 binding to ScH40 and CtH40. 110
Figure 4.11: Semi-analytical SEC of ScNob1 and DS0 RNA. 112
Figure 4.12: Semi-analytical SEC of ScNob1 and DS1. 113
Figure 4.13: Initial standard EMSA using ScNob1 and DS0 or DS1 RNA. 114
Figure 4.14: PAGE analysis of ScNob1 without and with DS0 RNA using different EMSA buffer systems. 115
Figure 4.15: Optimised EMSA using ScNob1 and DS0 or DS1. 115
Figure 4.16: EMSA results for DS0-DS10 RNAs with ScNob1. 117
Figure 4.17: ITC results for DS0 and DS1 RNAs binding to ScNob1. 120
Figure 4.18: ITC results for the DS6 RNA and the 6.1-5 mutants..... 122
Figure 4.19: ITC results for mutants DS6.6-11..... 123

Figure 4.20: ITC results for DS11, DS12, DScram1-2, and the human DS6 sequence with ScNob1. 124
Figure 4.21: Native gels of the RNAs used in ITC. 126
Figure 4.22: ITC results for DS6.5.2 binding to ScNob1. 127
Figure 4.23: Example EM images of ScNob1 alone and with a 2 times excess of DS1 RNA. 128
Figure 4.24: Semi-analytical SEC of ScNob1 and human DS1 RNA. 129
Figure 4.25: ITC results for the ITS1 fragment with ScNob1. 130
Figure 4.26: Semi-analytical SEC of ScNob1 and ITS1 RNA. 131
Figure 4.27: Cleavage assays using ScNob1 and DS5 RNA. 133
Figure 4.28: EMSA (top) and cleavage assay (bottom) comparing the WT ScNob1 and the D15N mutant. 134
Figure 4.29: Mini Q anion exchange purification of *in vitro* transcribed DS1. 136
Figure 4.30: Semi-analytical SEC results using ScNob1 with each of the DS1 peaks obtained from the Mini Q column. 137
Figure 4.31: Co-purification SEC results for ScNob1 and DS1. 138
Figure 5.1: Dim2 sequence alignment using two archaeal sequences and the human and yeast proteins. 149
Figure 5.2: Structural comparison of the *P. horikoshii* (3AEV) structure (PhDim2) to the ScDim2 model generated by Phyre. 150
Figure 5.3: Model validation using side chain and backbone geometry. 151
Figure 5.4: Expression trials of full-length (FL) ScDim2 constructs present prior to the start of this project. 152
Figure 5.5: Expression trials for the different ScDim2 constructs created for this project. 154
Figure 5.6: Expression trials of the human Dim2 (HsDim2) constructs. 155
Figure 5.7: Expression trials of the codon optimised CtDim2 in 2xTY medium. 155
Figure 5.8: Purification trials for ScDim2. 157
Figure 5.9: Nickel affinity purification of His-tagged ScDim2. 158
Figure 5.10: Purification of the GST-tagged HsDim2. 159
Figure 5.11: Purification of the His-tagged ScDim2ΔN51 protein. 161
Figure 5.12: Purification of full-length GST-tagged ScDim2 using the protocol established for the human protein. 162
Figure 5.13: HsDim2 and ScDim2ΔN51 SEC results compared to size standards. 163
Figure 5.14: Thermal denaturation results for HsDim2. 165
Figure 5.15: Limited proteolysis of HsDim2. 166
Figure 5.16: Proteolysis of the HsDim2 protein over time using 1:100 Elastase or GluC. 167

Figure 5.17: MALDI-ToF analysis results for the full-length and proteolysed HsDim2. 168
Figure 5.18: Preparative proteolysis and SEC purification of the HsDim2 protein. 169
Figure 5.19: Semi-analytical SEC results using ScNob1 and ScDim2ΔN51. 170
Figure 5.20: Semi-analytical SEC of ScNob1 and ScDim2ΔN51 at high pH (8.5 instead of 7.5). 171
Figure 5.21: Semi-analytical SEC analysis of ScNob1 binding to ScDim2ΔN51 at higher salt concentrations (350 mM instead of 150 mM KCl). 172
Figure 5.22: Semi-analytical SEC results for ScNob1 mixed with HsDim2. 174
Figure 5.23: Semi-analytical SEC using ScDim2ΔN51 and ScDim1. 175
Figure 5.24: Semi-analytical SEC of ScNob1 and ScDim1. 176
Figure 5.25: Semi-analytical SEC of ScNob1 and ScFap7. 177
Figure 5.26: Semi-analytical SEC using ScDim2ΔN51 and yeast DS0 RNA. 178
Figure 5.27: Semi-analytical SEC using HsDim2 and yeast DS1 RNA. 179
Figure 5.28: Semi-analytical SEC using ScDim2ΔN51 and yeast DS1 RNA. 180
Figure 5.29: Semi-analytical SEC using HsDim2 and *in vitro* transcribed human DS1 RNA. 182
Figure 5.30: Semi-analytical SEC using ScDim2ΔN51 and yeast DS1 RNA at pH 8.5. 183
Figure 5.31: Semi-analytical SEC using ScDim2ΔN51 and yeast DS1 RNA in 350 mM KCl. 184
Figure 5.32: The raw sensogram results obtained from the SPR experiments at different temperatures. 185
Figure 5.33: The analysis results for the SPR experiment performed at 10˚C. 186
Figure 5.34: Semi-analytical SEC using HsDim2, ScNob1, and the yeast DS1 RNA. 188
Figure 6.1: Sample crystals obtained using ScTsr1Δ410-476. 197
Figure 6.2: Sequence alignment of the full-length yeast Tsr1 and the ΔN46Δ410-476 mutant. 199
Figure 6.3: Crystals obtained using ScTsr1ΔN46Δ410-476. 200
Figure 6.4: Silver-stained SDS-PAGE gel of the ScTsr1ΔN46Δ410-476 protein and the crystals shown in the previous figure. 201
Figure 6.5: Sequence alignment of ScTsr1 and archaeal eIF5B and SelB. 203
Figure 6.6: Purification of the seleno-methionine labelled ScTsr1ΔN46Δ410-476. 209
Figure 6.7: Crystals obtained with the seleno-methionine labelled ScTsr1ΔN46Δ410-476. 211
Figure 6.8: Fluorescence scan results for the seleno-methionine labelled ScTsr1ΔN46Δ410-476 crystals. 211

Figure 6.9: Overlay of the full-length and truncated archaeal eIF5B (PDB: 1G7R) and SelB (PDB: 4ACA). 214
Figure 6.10: A sample image of the electron density obtained using multi-wavelength anomalous dispersion (MAD) experimental phasing. 217
Figure 7.1: Model of the D site cleavage combining data obtained in this project and the translation-like cycle model previously described (see Figure 1.4). 221

i. Abbreviations

1M7	1-methyl-7-nitroisatoic anhydride
AEBSF	4-(2-Aminoethyl)benzenesulfonyl fluoride
AC	Ammonia/CAPS buffer
ASU	Asymmetric unit
ATP	Adenosine 5′ tri-phosphate
BIS-TRIS	2,2-Bis(hydroxymethyl)-2,2',2''-nitrilotriethanol
CAPS	*N*-cyclohexyl-3-aminopropanesulfonic acid
CHCA	α-Cyano-4-hydroxycinnamic acid
CRAC	Cross-linking and cDNA analysis
Ct	from *Chaetomium thermophilum*
C-terminus	Carboxyl terminus
dATP	Deoxy-adenosine 5′ tri-phosphate
DLS	Dynamic light scattering
DMS	Dimethyl sulfate
DNA	Deoxyribonucleic acid
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
dTTP	Deoxy-thymidine 5′ tri-phosphate
EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscopy
EMSA	Electrophoretic mobility shift assay
ETS	External transcribed spacer
FL	Full-length
FT	Flowthrough
GDP	Guanine 5′ di-phosphate
GSH	Glutathione
GST	Glutathione S-transferase
GTP	Guanine 5′ tri-phosphate
H40	Helix 40 RNA
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His tag	Hexahistidine tag
Hs	from *Homo sapiens*
HPLC	High-performance liquid chromatography
IF2	Initiation factor 2
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
ITS	Internal transcribed spacer
K_d	Dissociation constant

LIC	Ligase-independent cloning
LSU	Large subunit
MALDI-ToF	Matrix-assisted laser desorption/ionization – time-of-flight
MES	2-(*N*-morpholino)ethanesulfonic acid
MPD	2-Methyl-2,4-pentanediol
MR	Molecular replacement
MW	Molecular weight
NAD	Nicotinamide adenine dinucleotide
Ni-NTA	Nickel-NTA
NMR	Nuclear magnetic resonance
NTA	Nitrilotriacetic acid
N-terminus	Amino terminus
OD	Optical density
PCR	Polymerase chain reaction
PDB	Protein databank
PEG	Polyethylene glycol
Ph	from *Pyrococcus horikoshii*
pI	Isoelectric point
PNK	T4 Polynucleotide Kinase
R_g	Radius of gyration
R_h	Hydrodynamic radius
RMSD	Root-mean-square deviation
RNA	Ribonucleic acid
rpm	Revolutions per minute
rRNA	Ribosomal RNA
RT-PCR	Reverse transcriptase PCR
SAM	*S*-Adenosyl methionine
Sc	from *Saccharomyces cerevisiae*
SDS	Sodium dodecyl sulfate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SEC-MALS	SEC coupled to multi-angle light scattering
snoRNA	Small nucleolar RNA
SOC medium	Super Optimal Broth medium
SPR	Surface plasmon resonance
SSU	Small subunit
TBE	Tris/borate/EDTA buffer
TC	Tris/CAPS buffer
TCEP	Tris(2-carboxyethyl)phosphine
TFA	Trifluoroacetic acid

T_m	Melting temperature
tRNA	Transfer RNA
UV	Ultra-violet
V_M	Matthews coefficient
v/v	Volume per volume
WT	Wild type
w/v	Weight per volume

ii. Lay Summary

The ribosome is a molecular machine responsible for making all of the proteins inside every cell. Its assembly, in turn, requires the action of multiple proteins. Nob1, Dim2, and Tsr1 are three proteins out of the ~200 factors involved in the process. They are needed in order for the ribosome to become functional. This project aimed to explore the mechanism by which these three proteins are able to give rise to competent ribosomes.

Nob1 and Dim2 interact with each other and the immature ribosome. When these proteins are removed from cells, the ribosomes produced are unable to make proteins. Different parts of the ribosome have been shown to interact with the two proteins. However, here we show that Nob1 and Dim2 show an ability to bind to the same part. Thus, it is possible that they compete for this site. This could prove to be a very important part in regulating the ribosome production.

Tsr1 function is unknown to date. This protein has also been shown to bind to the ribosome and, without it, the cells are unable to make ribosomes capable of making proteins. This project aimed to characterise Tsr1 by obtaining its structure. The structure of a protein can often give ideas about its function. Some preliminary steps have been taken although the structure has not yet been obtained.

Overall, the research presented here clarifies certain aspects of ribosome production and forms a foundation for future studies.

iii. Abstract

Ribosome biosynthesis in eukaryotes is a complex process involving over 200 accessory factors. Nob1, Dim2, and Tsr1 are three conserved factors that are all involved in the late processing steps of the small subunit (40S) pre-rRNA. Depletion of any of these factors leads to the accumulation of the immature 20S pre-rRNA.

Nob1, an essential protein in yeast, performs the final cleavage of small subunit rRNA giving rise to the mature particle. It is aided in this process by other proteins such as Dim2. Previously, the two proteins have been shown to interact. Nob1 function was found to be more efficient in the presence of Dim2.

Previous studies also indicated that Nob1 binds a site on the pre-40S that is distal to the cleavage site while Dim2 binds proximally. Using analytical gel filtration, electrophoretic mobility shift assays, and isothermal titration calorimetry we show that Nob1 does not interact with the distal binding site *in vitro*. Instead, a stable complex with a micromolar disassociation constant can be formed with a sequence derived from the cleavage site. Thus, Nob1 and Dim2 appear to be competing for this site. The interaction with both proteins is blocked when this sequence is sequestered in a hairpin structure, which has been previously predicted to form at this site. By altering individual bases in the RNA sequence, we have identified the sequence determinants for Nob1-rRNA recognition.

Tsr1 function is unknown to date. It shares sequence similarity with certain GTPases; however, no GTP binding has been identified in previous studies. The depletion of this factor leads to a similar phenotype as the depletion of Nob1 and Dim2. By screening various deletion constructs, we have obtained good quality, diffracting crystals of yeast Tsr1. However, due to time constraints, the full structure has not been solved. Here we present the initial analysis of the crystallographic data and the potential for solving the structure in the future.

Overall, the data presented in this thesis bring insight into the final step of small subunit ribosome maturation.

1 INTRODUCTION

1.1 Ribosomal Function and Abnormalities

Ribosome function is essential for life in all organisms. These macromolecular machines are needed in every cell in order to synthesize protein chains using the peptidyl transferase reaction (Green & Noller, 1997). It has been known for a long time that ribosomes are comprised of proteins and RNA and that the two components come together to perform the catalytic reaction, thus making these molecules ribozymes (Green & Noller, 1997; Nissen et al, 2000; Noller et al, 1992).

Understanding how these molecules are made and how they are regulated is important for multiple aspects of biological and medical sciences. In the past, it was thought that any defect in the ribosome biosynthesis pathway would be lethal. More recently, abnormalities in the ribosome have been implicated in the pathogenesis of a number of diseases. For example, Diamond-Blackfan anaemia (autosomal dominant) and Shwachman-Bodian-Diamond syndrome (autosomal recessive) are rare disorders caused by mutations in ribosomal proteins (Narla & Ebert, 2010). Other diseases, such as Bowen-Conradi syndrome (autosomal recessive) and Treacher Collins syndrome (autosomal dominant), are caused by mutations in the ribosomal RNA (rRNA) processing proteins (Freed et al, 2010).

These disorders are caused by dysregulation of the ribosome biogenesis pathway, which, in turn, causes the accumulation of the proteins involved. This leads to the activation of the p53 stress response by inhibition of its E3 ubiquitin ligase, Mdm2. Haploinsufficiencies are often enough to activate the pathway. The p53 response can range from arresting the cell cycle to apoptosis depending on the severity of the defect. Thus, the cells are forced to evolve to survive and to supress the stress response leading to severe phenotypes such as the disorders mentioned. Also, it increases the risk of tumourgenesis as the p53 response suppresses uncontrolled proliferation. Thus, ribosome dysfunction is also associated with cancer (Bursac et al, 2014; Golomb et al, 2014).

In cancer, cells need higher levels of protein production during proliferation. Recent studies showed that knockdown of a ribosome assembly factor, Nob1, leads to decreased proliferation and tumour genesis in ovarian cancer (Lin et al, 2011), hepatocellular carcinoma cell lines (Lu et al, 2011), and in *in vitro* and *in vivo* models of non-small cell lung cancer (Li et al, 2014). The mechanism for this effect remains unknown. In malignant glioma, Nob1 has been suggested to potentially be an oncogene. Its levels in cell lines and in glioma patients were correlated to the severity of the disease (Zhou et al, 2013). All of these findings make Nob1 a potential anti-cancer target and thus, make it an important target for study.

Understanding ribosome biogenesis in organisms other than humans is also important. This process in pathogens such a trypanosomatids, (e.g. *Trypanosoma brucei*) requires additional factors (Jensen et al, 2005; Prohaska & Williams, 2009). These are not present in metazoa and can potentially be used as drug targets against these pathogens. Understanding the pathway in both types of organisms is crucial for developing such drugs.

1.2 Overview of the Early rRNA Processing in Eukaryotes

Ribosome assembly in eukaryotes is a complicated, multi-step process involving over 200 accessory proteins and multiple cellular compartments (Henras et al, 2008). It is more fully understood in prokaryotes where it is simpler and involves fewer proteins (Connolly & Culver, 2009; Hage & Tollervey, 2004). Also, the bacterial ribosome has been fully assembled *in vitro* (Traub & Nomura, 1968). In eukaryotes, the process is largely conserved and has been extensively studied in the model yeast, *Saccharomyces cerevisiae* (Kressler et al, 2010). However, while the majority of factors and some processes involved have been uncovered, the mechanistic details of how the factors work together to bring about the necessary maturation events remain largely unknown.

The assembly process starts during rRNA transcription in the nucleolus. The 5S rRNA is transcribed by RNA polymerase III while the other three out of the four rRNAs (18S, 5.8S, and 25S) are produced as a single transcript (35S) by RNA polymerase I (Kressler et al, 2010). During transcription, this large rRNA molecule

interacts with multiple proteins to give a 90S pre-ribosome (Schafer et al, 2003). Then, multiple cleavage steps occur in order to separate and mature the individual components (Kressler et al, 2010).

The cleavages of yeast and human rRNA are summarised in **Figure 1.1**. In yeast, cleavages A_0 and A_1 remove the 5′ external transcribed spacer (ETS). Then, cleavage at site A_2 in the internal transcribed spacer (ITS) separates the small subunit (SSU, pre-40S) and large subunit (LSU, pre-60S) pre-rRNAs. This cleavage occurs co-transcriptionally, while the LSU rRNA is still being made (although it can also take place after transcription, depending on the metabolic state of the cell) (Kos & Tollervey, 2010; Osheim et al, 2004). The processing in the nucleolus involves multiple small nucleolar RNAs (snoRNAs) and associated proteins (Dragon et al, 2002). The pre-40S particle is then released and rapidly transported into the cytoplasm for further processing (Kressler et al, 2010).

After the LSU rRNA is transcribed it is also released. The LSU rRNA travels from the nucleolus to the nucleoplasm and eventually to the cytoplasm with different processing steps occurring in the different compartments, while the SSU rRNA does not appear to undergo processing in the nucleoplasm (Kressler et al, 2010).

In human cells, the processing steps are very similar but a different numbering system for the cleavages is used, as seen in **Figure 1.1**. The largest difference can be seen in the processing of the SSU rRNA. After separation from the other RNAs by cleavage at site 2 (A_2 in yeast), a further cleavage has to take place prior to export into the cytoplasm. This is known as cleavage E. Otherwise, the process is very similar although other small differences do exist (Carron et al, 2011; Rouquette et al, 2005). For example, while the yeast polycistronic transcript is processed co-transcriptionally, the human one is not (Turowski & Tollervey, 2014). This is likely due to the relative growth rates of the cells. This research concentrates on the late maturation of the 18S rRNA particle.

1.3 Overview of the Late 18S rRNA Processing in Eukaryotes

The SSU pre-rRNA is rapidly exported from the nucleolus to the cytoplasm where two main processing steps occur that allow the mature 18S rRNA to be

created. Firstly, the RNA becomes dimethylated at two 3′ adenines (1781 and 1782) by Dim1 (Campbell & Karbstein, 2011). Although, this modification does not appear to be necessary for processing *in vivo*, cleavage at the D site cannot occur without the dimethylation *in vitro* (Kos & Tollervey, 2010). Finally, cleavage at site D (in yeast) is needed to mature the 20S pre-rRNA into the 18S rRNA (Campbell & Karbstein, 2011; Hage & Tollervey, 2004; Kressler et al, 2010; Lamanna & Karbstein, 2011).

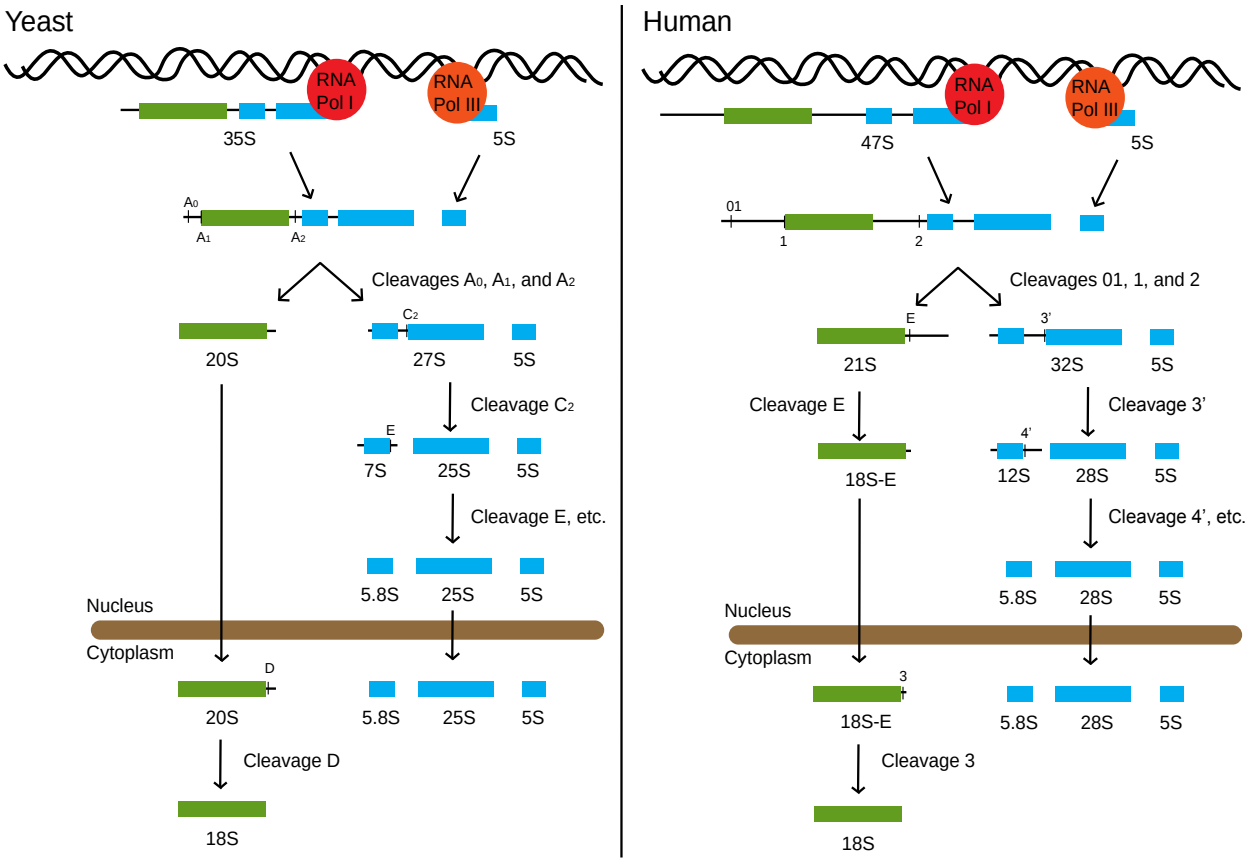


Figure 1.1: Summary of the cleavage steps involved in making the mature rRNAs in *S. cerevisiae* and *H. sapiens*.

The small subunit (SSU) rRNA is shown in green while the large subunit (LSU) rRNA is shown in blue. Cleavages at sites A_0 and A_1 (or 01 and 1 for human cells) remove the external transcribed spacer at the 5′ end. Cleavage A_2 (or 2) separates the small and large subunit rRNAs. The LSU rRNA undergoes multiple maturation steps in the nucleoplasm and is separated into its constituent parts (by cleavage at site C_2 or 3') before export into the cytoplasm. In the cytoplasm, final maturation occurs. The yeast SSU rRNA, in contrast, is not processed in the nucleoplasm and is rapidly exported into the cytoplasm where it undergoes the final cleavage at site D to give the mature 18S fragment. The human SSU rRNA undergoes an additional cleavage in the nucleus at site E before the export and final cleavage at site 3. RNA polymerase III transcribes the 5S rRNA separately from the other components.

Multiple factors remain bound to the nucleic acid throughout the transport into the cytoplasm. In *S. cerevisiae*, these include Rio2, Tsr1, Enp1, Ltv1, Dim1, Dim2 (also known as Pno1), and Nob1, all of which are involved in the final processing steps (Campbell & Karbstein, 2011; Fatica et al, 2003; Schafer et al, 2003; Vanrobays et al, 2003; Vanrobays et al, 2008). Homologous proteins have been shown to bind to the pre-40S particle in human cells (Wild et al, 2010; Zemp et al, 2009) implying that late processing steps are conserved. Rio2, Ltv1, and Dim2 have all been shown to contain functional nuclear export signals (NES) that aid in the export of the premature particle (Seiser et al, 2006; Vanrobays et al, 2008; Zemp et al, 2009).

The seven accessory factors appear to make two distinct modules in yeast. The first module consists of Enp1 and Ltv1, which have been shown to form a stable complex, along with Rps3 (a ribosomal protein), that is resistant to high salt (Schafer et al, 2006). The other module consists of Tsr1, Dim1, Dim2, and Nob1, which are all connected through their interaction with Rio2 (although the Tsr1 – Rio2 interaction has not been confirmed) (Campbell & Karbstein, 2011). Interactions between Dim2 and Dim1 or Nob1 have also been mapped (Campbell & Karbstein, 2011; Tone & Toh, 2002; Vanrobays et al, 2004) although the former might require an RNA mediator (Campbell & Karbstein, 2011). Previously mapped interaction between these late acting accessory factors are shown in **Figure 1.2**. The interactions shown here are strong and appear to be direct (Campbell & Karbstein, 2011).

The Enp1/Ltv1 module is required for earlier processing (A_0, A_1, and A_2 cleavages) as well as at the late stages (Chen et al, 2003). In late processing, depletion of Ltv1 or mutation of its NES results in 20S pre-rRNA accumulation (Fassio et al, 2010; Lebaron et al, 2009). This could be due to interference with the Prp43 helicase activity (Lebaron et al, 2009) or due to other unknown effects (Fassio et al, 2010). Prp43 is an RNA helicase, which has been shown to be needed for the cleavage at site D (Pertschy et al, 2009). Pfa1 is a protein, which stimulates Prp43 activity. Deletion of Ltv1 and Pfa1 gives a more dramatic inhibition of the 20S pre-rRNA processing (Lebaron et al, 2009) implying that the two might be connected.

The second module has a more direct role. As mentioned earlier, Dim1 methylates the RNA (Campbell & Karbstein, 2011). Nob1 is the endonuclease

responsible for cleavage at site D (Pertschy et al, 2009). A mutation that abolishes just its nuclease activity (D92N) prevents the cleavage from happening (Fatica et al, 2003; Fatica et al, 2004). Nob1 has also been shown to bind near the D site (Lamanna & Karbstein, 2009). The other proteins in this module are also required for the cleavage to occur. Depletion of Dim2 prevents both late steps: the methylation by Dim1 and cleavage by Nob1 (Woolls et al, 2011); whereas, the kinase activity of Rio2 has been shown to be required for the D site cleavage and for the release of Nob1, Dim2, and Ltv1 from the mature particle (Ferreira-Cerca et al, 2012; Zemp et al, 2009). Tsr1 is also involved in late 20S rRNA processing although its exact function is unknown. Depletion of this protein causes the processing to slow down and the 20S pre-rRNA to accumulate (Gelperin et al, 2001; Strunk et al, 2011).

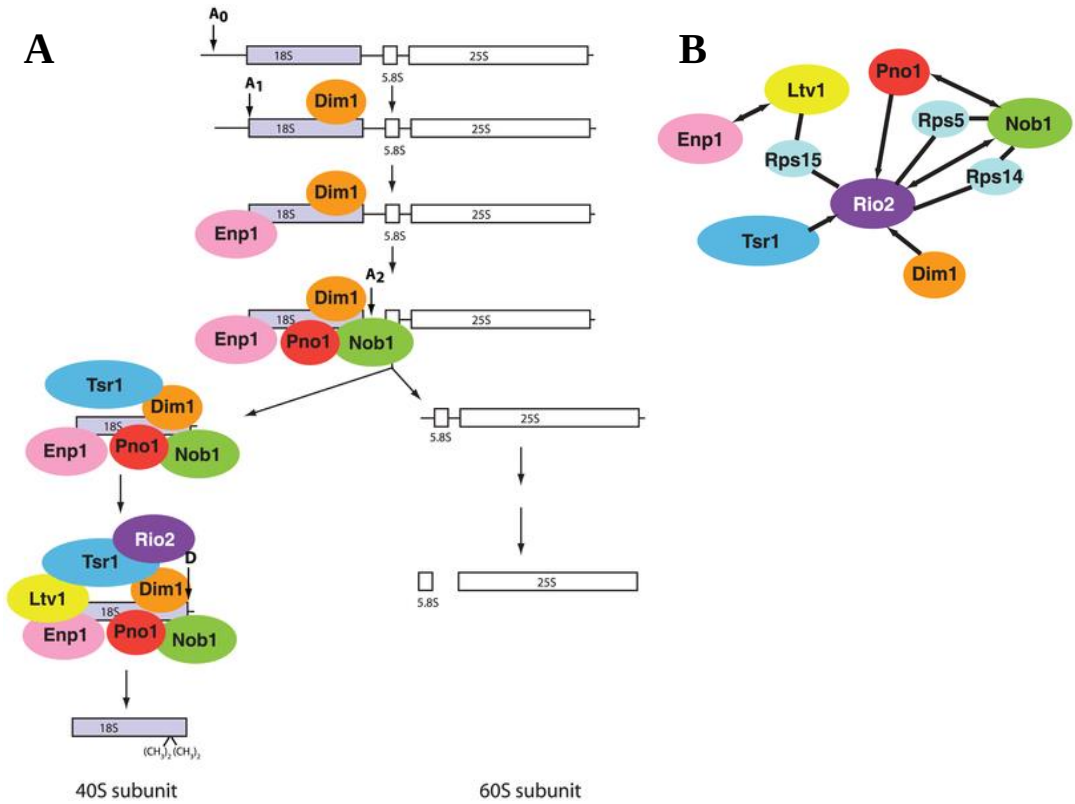


Figure 1.2 (Campbell & Karbstein, 2011): Some main assembly factors involved in SSU rRNA processing.

A) Factors interacting with the SSU rRNA during different stages of maturation. The interactions between these factors have been mapped (see below). Multiple factors remain bound to the pre-40S particle throughout the nuclear export process. For simplicity, no ribosomal proteins are shown.

B) Map of the interactions observed by pull-down experiments (i.e. strong interactions) including the interaction between Nob1 and Pno1 (i.e. Dim2). Some ribosomal proteins (light blue) are included.

Dim1 is the most conserved out of all seven co-factors as its homolog can even be found in bacteria (known as KsgA) (Lafontaine et al, 1994; O'Farrell et al, 2006). The structure of KsgA has been solved to 2.1 Å, showing the structural mechanism involved in the methylation reaction (O'Farrell et al, 2004). Structural information for archaeal (LaRonde-LeBlanc & Wlodawer, 2004) and eukaryotic (Ferreira-Cerca et al, 2012) Rio2 has been published and clarified that the protein acts as an ATPase as well as a kinase. Finally a crystal structure of archaeal Dim2 (Jia et al, 2010; Jia et al, 2007) and a limited NMR model of archaeal Nob1 have been solved (Veith et al, 2012). Not much is known about the domain organisation of the eukaryotic accessory proteins.

The positioning of the seven late binding accessory factors on the premature 40S ribosome was determined using cryo-electron microscopy (cryo-EM) in combination with sequential depletion of the proteins to identify their densities (Strunk et al, 2011). **Figure 1.3** shows the results of the study.

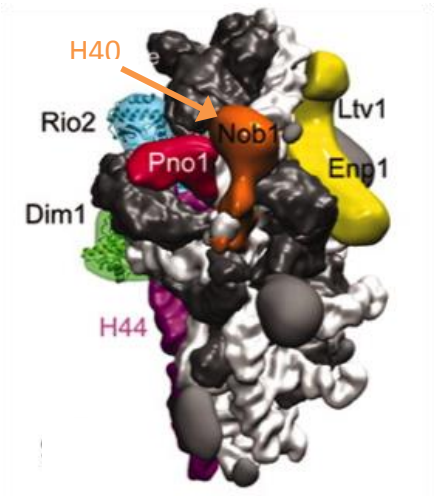


Figure 1.3 (Strunk et al, 2011): Positioning of some of the assembly factors on the pre-40S ribosome as shown by EM at 18 Å resolution.
Tsr1 is on the other side of the particle. Pno1 (also known as Dim2) and Nob1 are shown to be close to each other and close to the end of helix 44 where the D site cleavage occurs (shown in purple) and to helix 40 (marked with an arrow). Enp1 and Ltv1 can be seen to form the separate module. Solved structure of Rio2 and Dim1 can be docked into the densities.

Proteins, which are not structural components of the pre-ribosomes, are also important in the D site cleavage. As mentioned earlier, Prp43 activity is required to give the mature particles (Lebaron et al, 2009). Rio1 kinase is also needed as its activity allows the recycling of Nob1 and Dim2 following the cleavage (Widmann et al, 2012). Another example of this can be seen with Fap7. Fap7 adenylate kinase depletion leads to accumulation of the 20S pre-rRNA. Its interaction with the pre-40S particle appears to be transient but important (Granneman et al, 2005). A recent

study has shown that its activity is required for positioning of Rps14 and making the pre-ribosome a viable substrate for D site cleavage (Loc'h et al, 2014).

Finally, all of the seven accessory factors also prevent premature activation of the pre-40S particle. Tsr1, Rio2, and Dim1 binding sites (as shown by cryo-electron microscopy, or cryo-EM) overlap the sites required for binding of translation initiation factors eIF1 and eIF1A while Nob1 and Dim2 positioning is incompatible with eIF3 binding. The Enp1/Ltv1 module overlaps the hinge region so that the mRNA channel cannot open even if a translation initiation factor was to bind (Strunk et al, 2011).

This study aims to clarify the process of the 20S pre-rRNA maturation by characterising three biogenesis factors from eukaryotes: Nob1, Dim2, and Tsr1.

1.4 Previous Studies Performed on Nob1

Nob1 has been shown to be involved in the assembly of the 26S proteasome (Tone & Toh, 2002) as well as for the assembly of the 40S ribosome subunit. The sequence alignment between yeast and human Nob1 (**Figure 1.4**) shows moderate conservation, especially around the functional domains (PIN and zinc ribbon). Yeast Nob1 is seen to localise to the nucleus and the cytoplasm (Fatica et al, 2003) although it does not appear to be involved in early rRNA processing. It has been suggested that, in ribosome biogenesis, its activity is inhibited until later stages of rRNA maturation (Granneman et al, 2010). During 40S ribosome assembly, Nob1 interacts with multiple proteins including ribosomal proteins Rps5 and Rps14 (Campbell & Karbstein, 2011), and accessory factors such as Dim2 (Campbell & Karbstein, 2011; Tone & Toh, 2002; Woolls et al, 2011).

There is some evidence that Nob1 can exist as a tetramer although it is unclear if this is the case inside the cell. When the yeast protein was purified from bacteria as a recombinant construct, it appeared to form larger particles (i.e. tetramers) (Lamanna & Karbstein, 2009). However, the EM structure of the pre-40S ribosome shown earlier supports a Nob1 monomer binding the particle. A tetramer would not fit in the density observed.

Nob1 contains a PIN domain, which is required for D site cleavage (Fatica et al, 2004; Lamanna & Karbstein, 2009). Normally, this domain is associated with Mg^{2+}-dependent exonucleases (Arcus et al, 2004). *In vitro* experiments with the yeast protein showed that the nuclease activity of Nob1 is stimulated by manganese cations (Lebaron et al, 2012; Pertschy et al, 2009).

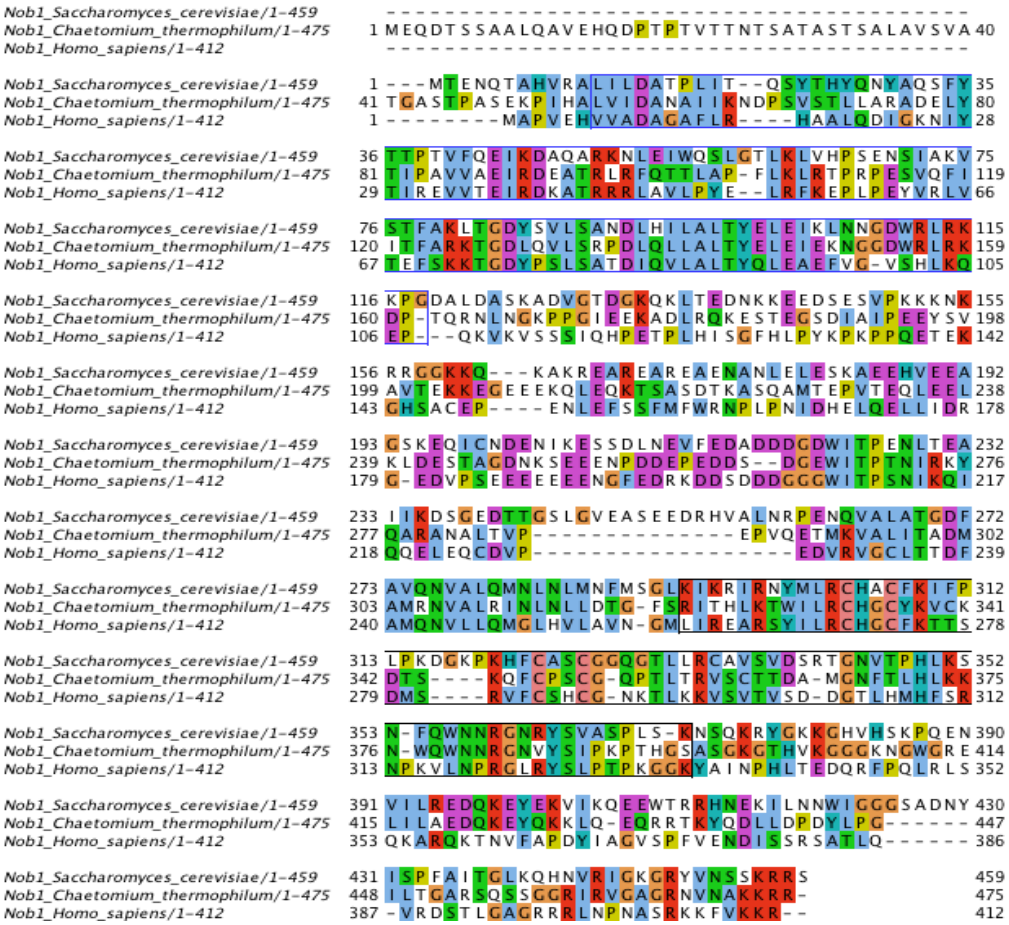


Figure 1.4: Sequence alignment of Nob1 from different organisms (obtained using ClustalW and Jalview).

The sequences are coloured using the Clustal colour scheme. Sequences for *S. cerevisiae*, *C. thermophilum*, and human proteins are shown. The functional domains are marked in boxes with outlines (the first blue one corresponds to the PIN domain and the second black box shows the zinc ribbon domain). The two fungi show an identity of 34.0%, which is increased to 44.4 and 46.8% in the PIN and zinc ribbon domains respectively. Out of the fungal proteins, the *C. thermophilum* Nob1 is more similar to the human protein (32.0% vs. 27.5% identity).

A study using DMS probing suggested that the PIN domain binds the 20S pre-rRNA close to the D site (Lamanna & Karbstein, 2009). However, another study using UV-cross-linking mapped Nob1 to a site that is distant from site D (helix 40)

(Granneman et al, 2010). This site is potentially too distant to allow cleavage implying that a conformational change is required to bring the D site in contact with the Nob1 PIN domain (Granneman et al, 2010; Lamanna & Karbstein, 2011). Studies using the archaeal protein showed that the zinc ribbon domain could bind to helix 40 (Veith et al, 2012). Thus, it is possible that both interactions occur via different domains of the protein. This is supported by the EM structure of the pre-40S ribosome, which places Nob1 in a position where it would be able to bind to both locations (**Figure 1.3**). Nevertheless, it is unknown how Nob1 is inhibited and why it is bound to the rRNA early in the processing pathway.

Currently, there is one model for D site cleavage that has been supported by the Tollervey and Karbstein labs independently (shown in **Figure 1.5**). In this model, there is a 'translation-like' cycle that involves interactions of the pre-40S with the 60S subunit. Both labs have shown that this interaction is mediated by the GTPase Fun12 (also known as eIF5B) (Lebaron et al, 2012; Strunk et al, 2012). Once the cleavage occurs, the Rli1 ATPase separates the ribosomal subunits in order for the accessory factors to dissociate and for a proper translation cycle to occur. Rli1 performs the same termination function in mature ribosomes (Strunk et al, 2012).

It was also found that the GTPase activity of Fun12 is needed to stimulate D site cleavage. It is possible that GTP hydrolysis causes conformational changes that allow Nob1 to perform the cleavage. However, the process is more complicated as D site cleavage stimulation also depends on the ability of Fun12 to connect the two ribosomal subunits (Lebaron et al, 2012).

Interactions between Nob1 and Prp43 (Pertschy et al, 2009), and Dim2 (Vanrobays et al, 2008; Woolls et al, 2011) have been shown to promote the cleavage. The physical interaction between Nob1 and Dim2 has been previously demonstrated using a yeast-two-hybrid screen (Tone & Toh, 2002) and pull-down experiments (Woolls et al, 2011). It is possible that Dim2 aids in Nob1 making the correct contacts.

Recently, Nob1 has also been shown to interact with Fap7 adenylate kinase *in vitro* (Hellmich et al, 2013). The interaction with Fap7 blocked the cleavage at the D site (Hellmich et al, 2013). However, without Fap7 expression, the D site cleavage cannot occur *in vivo* (Granneman et al, 2005). It was also shown, that *S. cerevisiae*

Nob1 can be inactivated by a point mutation in the PIN domain (D15N) (Lebaron et al, 2012).

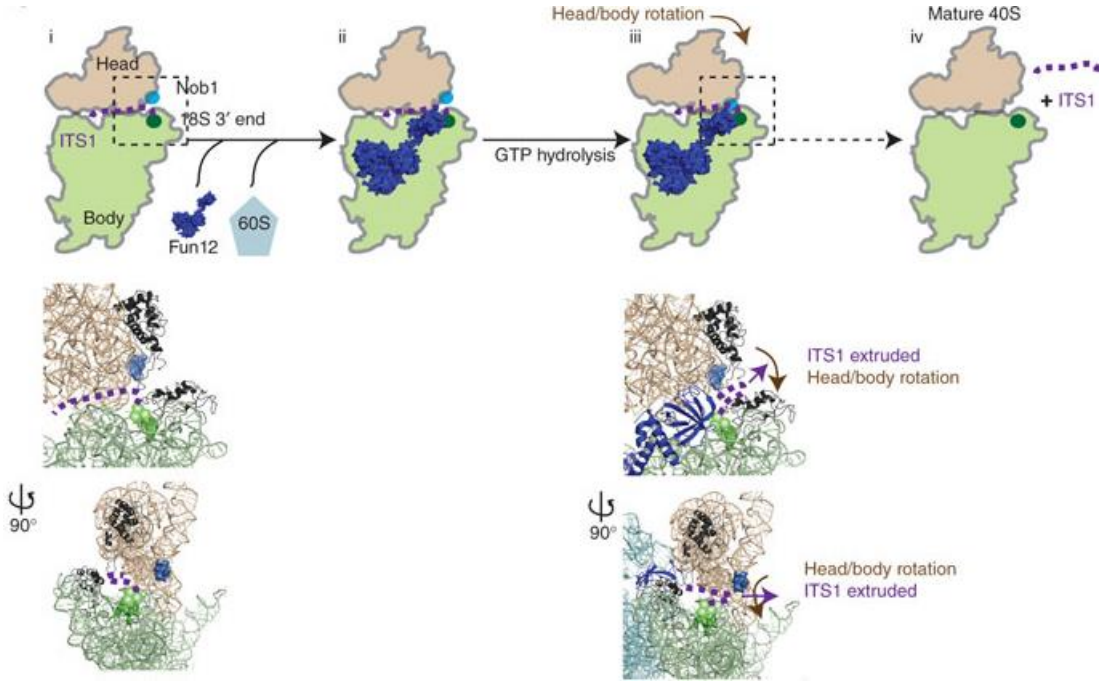


Figure 1.5 (Lebaron et al, 2012): Model for D site cleavage.
Only the pre-40S subunit is shown for clarity (it is divided into the brown head and green body domains). ITS1 is shown in purple, The Nob1 binding site is shown in blue, and the end of the mature 18S rRNA is shown in dark green. In the lower, structural images, locations of the ribosomal proteins Rps5 and Rps14 are shown in black. The ITS1 RNA is positioned inside the mRNA channel (i). Fun12 binds to the pre-40S along with the mature 60S subunit (ii). Fun12 hydrolyses GTP causing a string of conformational changes. The ITS1 is excluded from the mRNA channel. A potential head rotation occurs causing the D site to come into contact with Nob1. Thus, the D site cleavage can now proceed (iii). The accessory factors and the 60S subunit are released leaving a mature 40S particle (iv).

Finally, an NMR structure of archaeal Nob1 (from *Pyrococcus horikoshii*) has been published. **Figure 1.6** shows the structure of the two functional domains. The positioning of the domains with respect to each other could not be determined as they are connected by a flexible region and appear in different orientations in the structures (Veith et al, 2012). However, the data show how each of the domains in the protein is folded. Nevertheless, the archaeal protein displays a much shorter sequence than its eukaryotic counterpart (**Figure 1.7A**). While the archaeal protein structure gives a significant insight into Nob1 function, structural data on proteins from higher organisms are required to better understand its role in the context of the more complex eukaryotic ribosome.

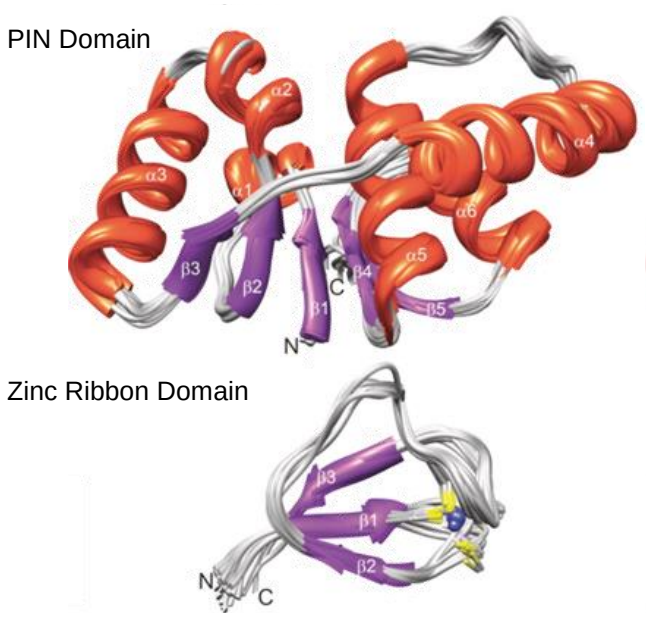


Figure 1.6 (Veith et al, 2011): The NMR-determined structures of the two functional domains of an archaeal (from *Pyrococcus horikoshii*) Nob1 protein (PDB ID: 2LCQ).
The top shows the PIN domain and the bottom shows the zinc ribbon domain. Each image is comprised of 20 overlaid structures. Alpha helices are shown in orange while the beta-strands are purple. The zinc divalent cation is shown as a blue sphere on the lower image. Four cysteine residues (sticks with sulphur atoms shown in yellow) coordinate its position.

1.5 Previous Studies Performed on Dim2

Dim2 is significantly conserved at the sequence level between yeast and humans (**Figure 1.8**), suggesting that its function is also conserved. Transgenic homozygous Dim2 knockout mice display early embryonic lethality. However, heterozygous knockout mice show normal development and fertility suggesting that a small amount of this protein is sufficient but necessary (Wang et al, 2012).

Like Nob1, Dim2 localises to the nucleus and the cytoplasm. However, unlike Nob1, Dim2 is needed for early processing (especially cleavage at site A_2) (Vanrobays et al, 2004). Crystal structures of an archaeal Dim2 (also from *P. horikoshii*) with (Jia et al, 2010) and without (Jia et al, 2007) rRNA bound to it have been published. Archaeal ribosome biogenesis and assembly are similar to eukaryotic ones although some key differences exist. For example, archaeal Dim2 contains two KH domains, which bind rRNA while the eukaryotic protein contains only one functional KH domain and a degenerated KH domain (Jia et al, 2010).

Thus, the archaeal structure could differ from the eukaryotic one. An alignment is shown in **Figure 1.7B** to highlight the differences.

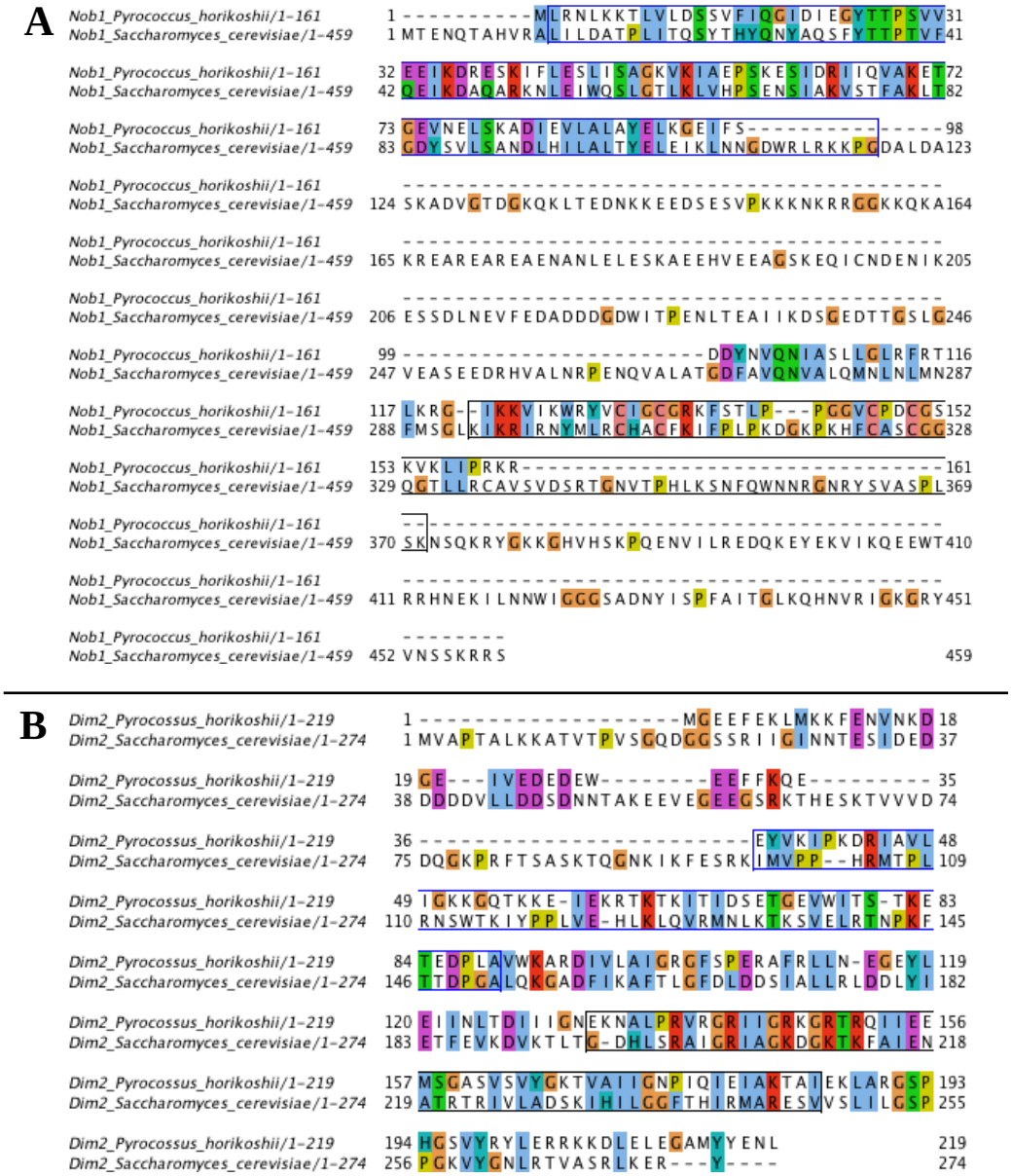


Figure 1.7: Sequence alignments of Nob1 and Dim2 from yeast and archaea (obtained using ClustalW and Jalview).

A) Nob1 alignment. The sequences are coloured using the Clustal colour scheme. Sequences for *S. cerevisiae* and *P. horikoshii* proteins are shown. The functional domains are marked in boxes with blue and black outlines as before. There is no significant similarity between the sequences except in the PIN domain (36.1% identity).

B) Dim2 alignment. Sequences are coloured in the same manner as the Nob1 alignment. The KH domain is shown in a black box. The KH-like domain is marked in a blue box (the first domain in the sequence). It forms a functional domain in the archaeal protein unlike in the yeast or human proteins. Both domains show some degree of conservation: 26.4% identity in the functional domain (black) and 22.6% in the degenerated one (blue).

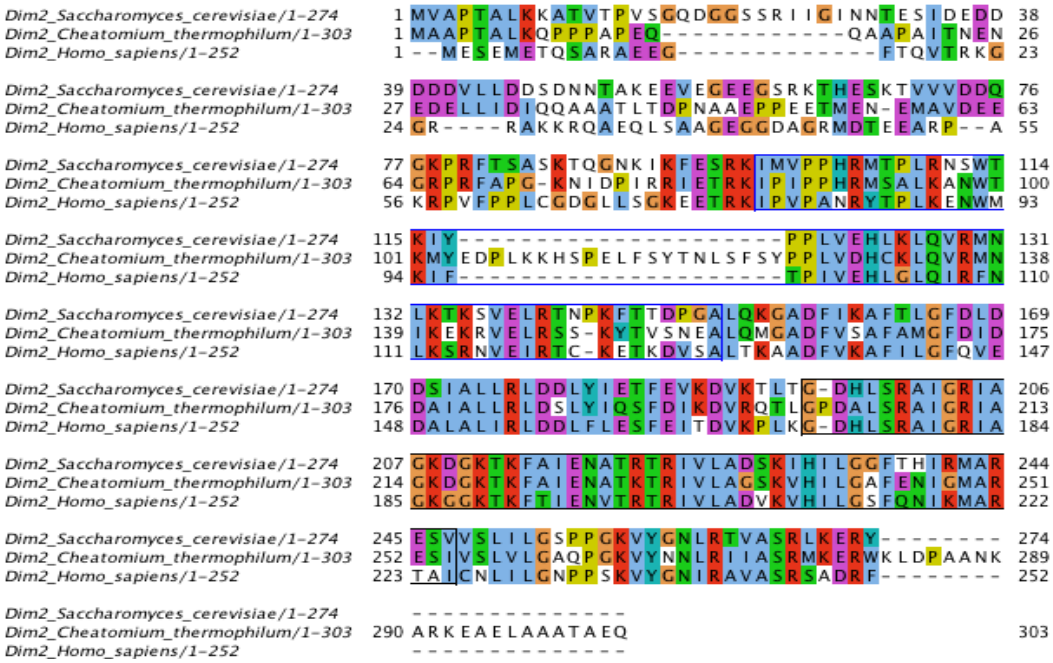


Figure 1.8: Sequence alignment of Dim2 from different organisms (obtained using ClustalW and Jalview).
Sequences are coloured in the same manner as the Nob1 alignment. The non-functional KH-like domain is shown in the first blue box and the functional KH domain is shown in the second black box. The two sequences from fungi show 49.66% identity, which is raised to 81.13% in the KH domain. In this instance, the *S. cerevisiae* protein is more similar to the human one (48.2% vs. 42.0% identity).

Nevertheless, the structure with rRNA allowed us to map the potential Dim2 binding site. **Figure 1.9** shows the general fold of the protein and the rRNA sequence derived from *Escherichia coli* that was copurified with archaeal Dim2. The binding site for the second KH domain (which is the functional domain in eukaryotes) is conserved from archaea to human. Thus, this is the predicted binding site in eukaryotes even though it has not been experimentally determined. The structure also showed that Dim2 interacts with initiation factor 2 (IF2) implying it might play a role in translation initiation (Jia et al, 2010).

Other studies have shown that Dim2 is necessary for the dimethylation reaction performed by Dim1 and the two proteins have been shown to co-precipitate (Vanrobays et al, 2004). Mutating a single residue in the functional KH-domain (G207A) is sufficient to prevent rRNA processing. The equivalent archaeal residue (G145) is in the second KH domain in a loop (marked with an arrow in **Figure 1.9A**) involved in making RNA contacts (Vanrobays et al, 2008). Dim2 is also needed for

pre-40S export from the nucleus and for efficient D site cleavage by Nob1 (Vanrobays et al, 2008; Woolls et al, 2011). It had been shown to be one of the last proteins to leave the maturing ribosome (Strunk et al, 2012). The release of Nob1 and Dim2 from the mature 18S particle seems to be dependent on the kinase activity of Rio2 (Zemp et al, 2009).

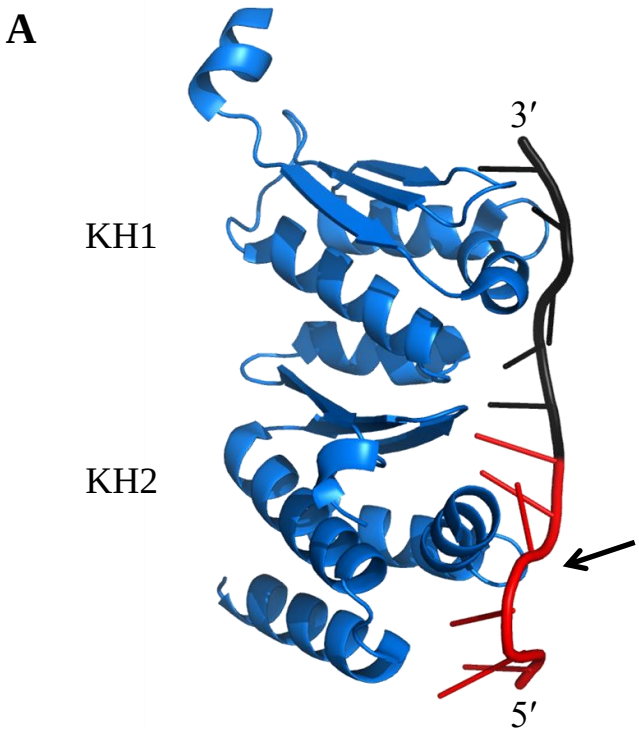


Figure 1.9: The general fold of archaeal Dim2 (blue) in complex with bacterial rRNA (black and red) (PDB: 3AEV).

A) The protein and nucleic acid are shown as cartoons. The two KH domains form the interactions with the nucleic acid. Both KH domains insert an α-helix into the space between adjacent bases forming multiple bonds to the backbone. The bases bend away from the helix and also form multiple non-covalent interactions with the protein. The binding appears to be sequence specific (CCUCC for KH1 and GGAUCA for KH2) although some changes in the sequence are tolerated (Jia et al, 2010). The loop contacting G145 is indicated with the arrow.

B) The sequences found near the 3′ end of SSU rRNA from archaea, yeast, and human. The preferred sequence needed for KH2 binding is conserved. Human and yeast Dim2 protein homologs only have the KH2 domain and lack the KH-1 domain. The rRNA sequences show that the mode of binding of KH2 is likely conserved.

Unlike Nob1, Dim2 shows some similarity to a prokaryotic protein, RbfA. In bacteria, this protein is required for 5′ processing of 16S rRNA. Like Dim2, it has a KH domain that is responsible for making the RNA-protein contacts (Datta et al, 2007; Strunk et al, 2011).

Figure 1.10 shows the secondary structure of the *S. cerevisiae* 20S rRNA. Cleavage site D is marked (blue circle), as is the mapped Nob1 binding site (red circle). The predicted site for Dim2 is shown in green. The position of these proteins on the three-dimensional structure of the ribosome particle was determined by electron microscopy (EM), as shown in **Figure 1.3** (Strunk et al, 2011).

The known interaction between Nob1 and Dim2, and the positioning of the two proteins on the rRNA molecule indicate that they are functionally related. Although Dim2 has been shown to aid in rRNA binding of Nob1 (Woolls et al, 2011), how the interaction of Nob1 with Dim2 mechanistically affects cleavage is unknown. The research described here aims to uncover the functional relationship between these proteins and D site cleavage.

1.6 Previous Studies Performed on Tsr1

The exact function of Tsr1 is unknown. The human and yeast proteins share about 30% sequence identity as seen in the alignment in **Figure 1.11**. It is highly similar to a GTP-binding protein known as Bms1. However, Tsr1 lacks the P-loop that is needed for GTP binding and the nuclear localisation signal (both of which are present in Bms1) (Gelperin et al, 2001). It also interacts with the pre-rRNA at a site very similar to that of the bacterial GTPase, Era (Granneman et al, 2010). In *E. coli*, Era is required for pre-30S maturation (Inoue et al, 2003). In yeast, depletion of Tsr1 leads to a decrease in 20S rRNA processing and an increase in 80S particles, as seen using cryo-electron microscopy (cryo-EM) (Gelperin et al, 2001; Strunk et al, 2011), showing that its role might be similar to that of Era in bacteria.

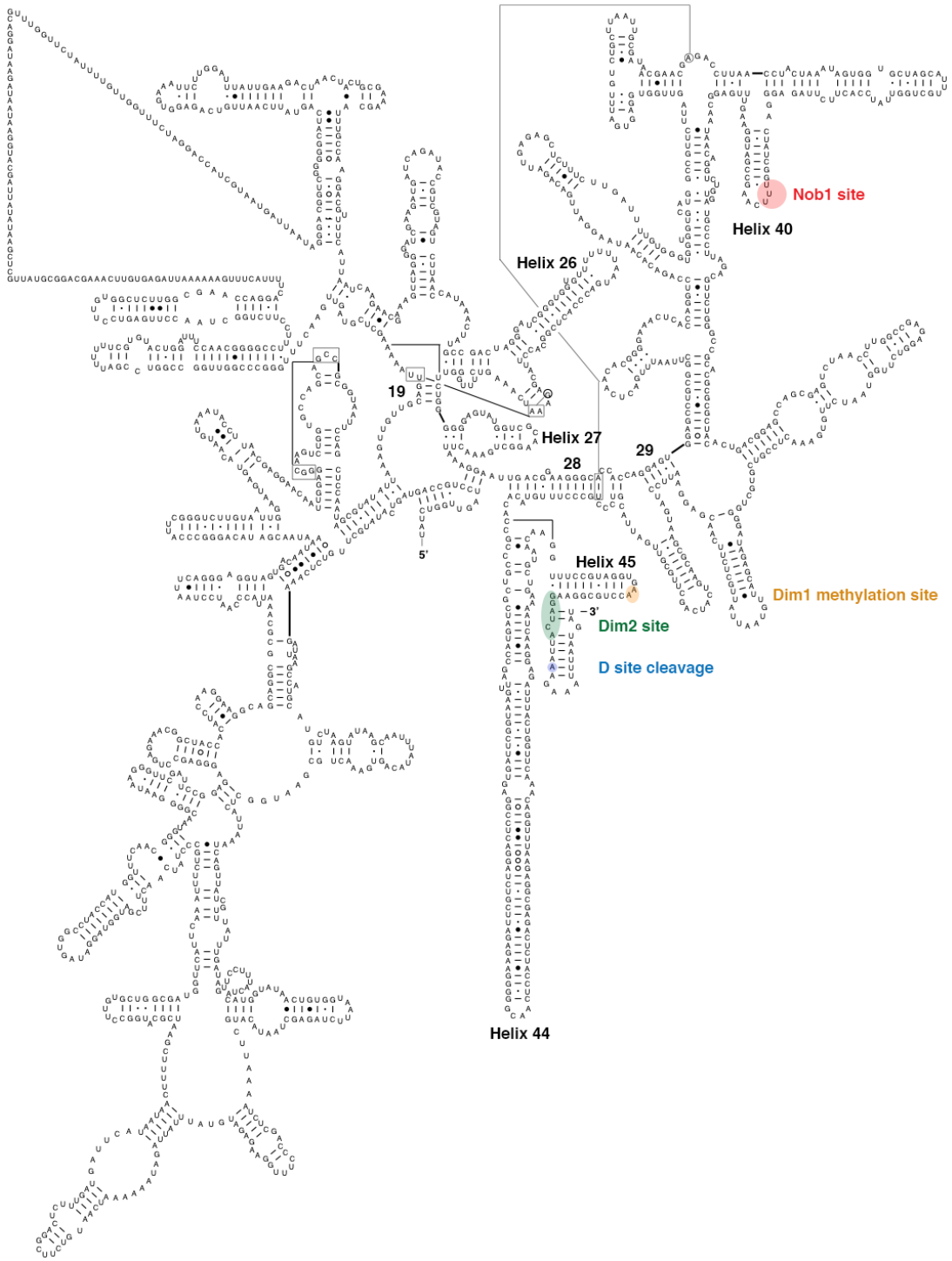


Figure 1.10: Secondary structure of the 20S rRNA from *S. cerevisiae* (obtained from http://www.rna.ccbb.utexas.edu/).

The 20S rRNA, apart from the extra ~200 bases on the 3′ end, is shown. Areas of interest for this project are marked. Helix 40 where Nob1 has been mapped to bind by cross-linking (red circle) is indicated. The D site cleavage takes place at the adenine marked with a blue circle. This site is distant from the Nob1 binding region. However, it very close to the Dim2 binding site (green oval) as predicted from the experiments performed on the archaeal protein. Helices of interest are marked with black numbers.

Tsr1 forms its major interactions with the central domain of the rRNA. In the folded structure, the protein is in close proximity to helices 44 and 45 and the D site although more contacts are made at helices 19 and 26-29 (labelled in **Figure 1.10**). There are 41 nucleotides at this location that are resistant to RNase digestion when Tsr1 is present confirming the binding position (Granneman et al, 2010). This positioning places the protein in a prime location to block premature ribosome assembly and translation initiation (Campbell & Karbstein, 2011; Granneman et al, 2010; Strunk et al, 2011). For example, the position of helix 28 of the rRNA prevents the binding of initiator tRNA. Also, the close proximity to the D site might impede premature cleavage by Nob1 (Granneman et al, 2010).

There is some evidence that Tsr1 prevents degradation of the premature 40S allowing the D site time to be cleaved. Also, the Tsr1 binding site overlaps with that of Rli1 ATPase. Therefore, its dissociation might be needed to allow Rli1 to bind and for the 40S to disassemble from the 80S-like ribosome formed in the 'translation-like' cycle that promotes D site cleavage (Strunk et al, 2012).

Finally, it has been suggested that Tsr1 is needed for nuclear export of the immature particles (Campbell & Karbstein, 2011; Gelperin et al, 2001; Schafer et al, 2003). However, as stated previously, its precise function and the mechanism by which it performs it are largely unknown. An atomic resolution structure of the protein might lead to an increased understanding of Tsr1.

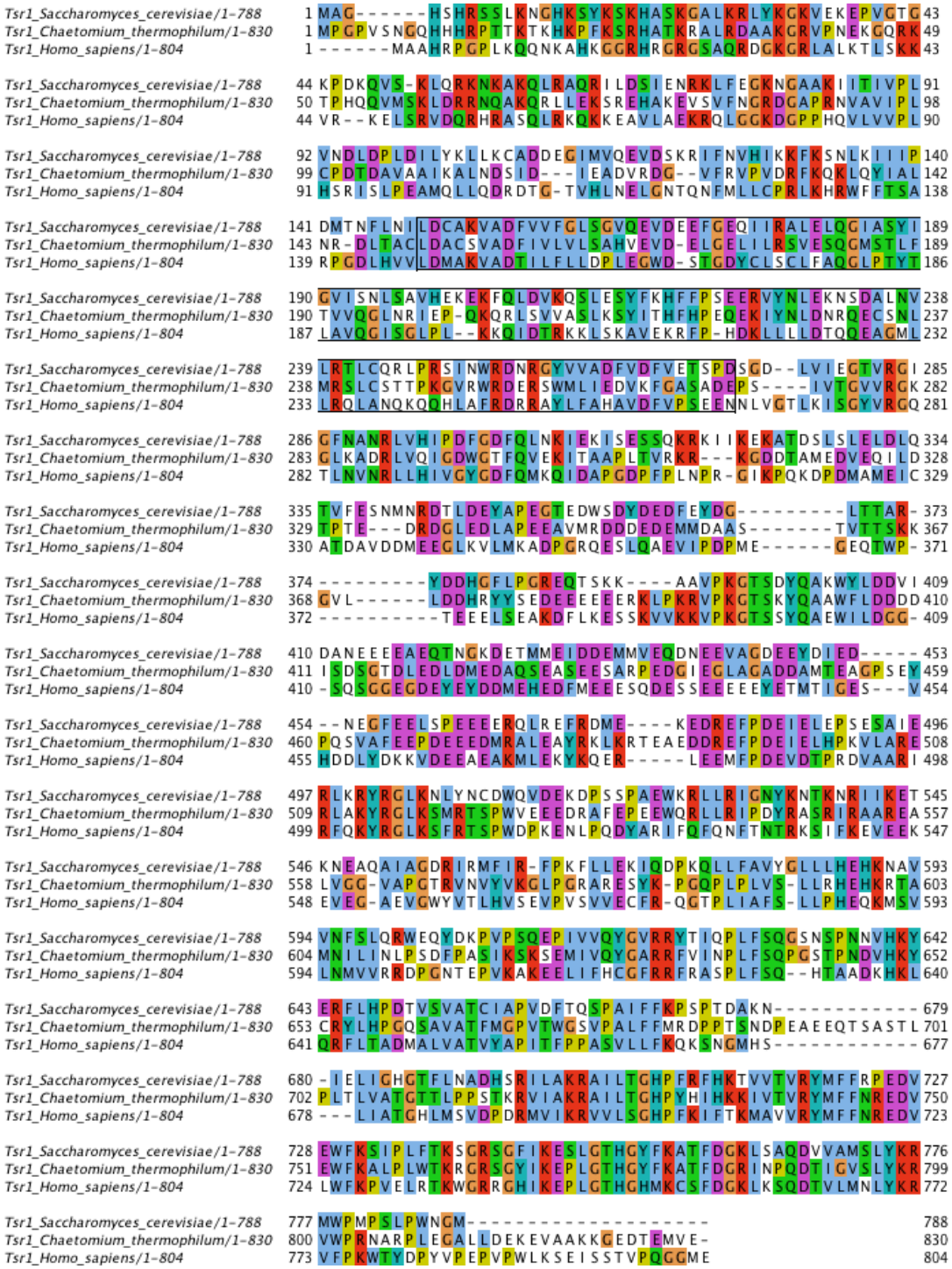


Figure 1.11: Sequence alignment of Tsr1 from different organisms (obtained using ClustalW and Jalview).

The sequences are coloured using the Clustal colour scheme. Sequences for *S. cerevisiae, C. thermophilum*, and human proteins are shown. The GTPase-like domain is marked in a black box. The two fungi show an identity of 37.8%, which is increased to only 38.7% in the GTPase-like domain. Out of the fungal proteins, the *S. cerevisiae* Tsr1 is more similar to the human protein (29.8% vs. 27.9% identity).

1.7 Model Organisms

The majority of the ribosome biosynthesis research performed to date utilizes *S. cerevisiae* as a model eukaryote (Kressler et al, 2010). Thus, the research undertaken in this study also aims to use this yeast as a model organism. The ribosome biogenesis pathway in eukaryotes is well conserved, although some differences do exist (Freed et al, 2010; Henras et al, 2008). The majority of total *S. cerevisiae* proteins are known to have orthologs or homologs (58%) in the overall human proteome (Karathia et al, 2011). Protein alignments shown in **Figures 1.4** and **1.8** show that Nob1 and Dim2 are reasonably well conserved, especially in the functional domains (PIN and zinc ribbon domains, and KH domain respectively). Structures of the yeast proteins would also supplement the previously performed research with a substantial background of information available.

The other organism chosen for studying the Nob1/Dim2 complex is *Chaetomium thermophilum*. This is a thermophilic fungus that has been found to survive in temperatures of up to 60˚C. Survival at such conditions implies that the proteins in the cell should be more stably folded. Thus, they make good candidates for structural studies. 73% of all known *S. cerevisiae* proteins have been found to have homologs in the *C. thermophilum* proteome (Amlacher et al, 2011). Nob1 and Dim2 are indeed well conserved as seen in the alignments in **Figures 1.4** and **1.8**.

Finally, attempts were also made to use the human proteins, as these are useful for understanding the pathology of human disorders.

1.8 Aims of Project

The general aim of the research presented in this thesis was to further the understanding of the late 20S processing in eukaryotic ribosome biosynthesis. There were two main questions that were investigated: how is the D site cleavage performed and what is the function of Tsr1 in the premature particle?

To answer these questions, the goals were to characterise the interaction between two eukaryotic proteins needed for D site cleavage, Nob1 and Dim2, and

RNA; also, to structurally characterise Tsr1 to increase the understanding of this protein's involvement in late 20S pre-rRNA maturation.

In order to achieve the first goal, initially, the four components needed for the D site cleavage (i.e. Nob1, Dim2, Nob1 RNA binding site, and Dim2 RNA binding site/D site) were to be obtained. This was to include expression and purification of *S. cerevisiae*, *C. thermophilum*, and *H. sapiens* proteins and transcription of the RNA binding sites. Once the proteins were obtained, trials would be performed for Nob1-Dim2 heterodimer *in vitro* reconstitution. These dimers and the individual proteins were then to undergo structural and biochemical characterisation with and without rRNA fragments. The effect of Dim2 on the Nob1-RNA interaction and on the D site cleavage was to be investigated.

For the second goal, a crystallisable, soluble Tsr1 construct was to be designed. The yeast protein was to be purified and crystallised. Then, the crystals were to be optimised and the structure was to be solved using experimental phasing.

The research presented in the next chapters shows the results of this investigation bringing some insight into the last steps of the small subunit maturation.

2 MATERIALS AND METHODS

1.1 Preliminary Cloning

Preliminary cloning was performed by Iva Tchasovnikarova and Dr. Atlanta Cook. The original yeast Nob1 construct in the pSL6 plasmid was obtained from the Tollervey lab. **Table 2.1** lists the clones already present prior to the start of this project. All the clones, including ones shown later, were tested for expression as described later.

Table 2.1: List of cloning performed prior to the start of this project.

Protein	Organism	Sequence	Plasmids	Tag
Nob1	*S. cerevisiae*	Full-length	pSL6	His
	S. cerevisiae	1-125 (PIN domain)	pEC-KHC, pEC-KGC	GST or His
	S. cerevisiae	264-459 (Zn ribbon)	pEC-KHC, pEC-KGC	GST or His
Dim2	*S. cerevisiae*	Full-length (FL)	pEC-KHC, pEC-KGC	GST or His
	S. cerevisiae	ΔN91	pEC-KGC	GST

1.2 Conventional Cloning

Table 2.2 contains a list of plasmids and inserts that were cloned using conventional cloning techniques. For this purpose, the gene of interest was amplified using PCR with primers containing designed restriction sites on each end. The engineered restriction sites can be also seen in **Table 2.2**. A list of all primers can be found in **Appendix I** while the PCR protocols are shown later, in standard laboratory procedures. Specific methods for the different clones can be seen below.

Table 2.2: Constructs created using conventional cloning.

Insert	Organism	Plasmid	5′ restriction site	3′ restriction site
Nob1 FL	*S. cerevisiae*	pET28α	NcoI	XhoI
20S rRNA transcription sequences	*S. cerevisiae*	pUC19	EcoRI	BamHI

1.2.1 ScNob1 cloning

The ScNob1 gene was amplified using the primers shown in **Appendix I** from a plasmid obtained from the Tollervey laboratory. The Phusion High Fidelity PCR Kit (NEB) was used for this purpose, according to manufacturer's instructions for a 50 µl reaction volume. The contents of the reaction were loaded onto a 1% w/v agarose gel containing 1 x SYBR Safe DNA Stain (Invitrogen) and run at 80V for 30 minutes. The band of the correct size (about 1377 bases) was excised and the gene was extracted using a Gel Extraction Kit (Qiagen).

The ScNob1 gene was then digested with NcoI and XhoI (NEB) in a 10 µl volume according to the manufacturer's instructions (using 10 units of each enzyme and 7 µl of the extracted DNA). The enzymes were inactivated by incubation at 80˚C for 20 minutes. 150 ng of the pET28α plasmid were also digested with 10 units of these enzymes in a 10 µl reaction. Again, the enzymes were inactivated at 80˚C. Following digestion, the plasmid was separated on a 1 % w/v agarose gel and extracted from the gel, as described previously for the ScNob1 gene.

The gene and the plasmid were ligated using T4 DNA ligase (NEB) in a 10 µl reaction volume. The manufacturer's instructions were followed. 2 µl of the gel purified vector and 6 µl of gene were used in the reaction. The mixture was incubated at room temperature for 30 minutes and transformed into XL-1 *E. coli* strain following a standard transformation protocol for chemically competent cells. Selection was performed based on kanamycin resistance.

Four colonies were picked and grown overnight in 3 ml LB supplemented with kanamycin. The plasmid was then extracted using a MiniPrep kit (Qiagen). The plasmids were digested with 5 units XhoI and BamHI (an extra site in insert) (NEB) according to manufacturer's instructions (using 3 µl of extracted DNA). Positive clones were sent for Sanger sequencing. All of the sequencing was performed by GenePool at the University of Edinburgh.

During the course of this project, Dr. Atlanta Cook also created five ScNob1 loop-out mutants in the pET28α plasmids. These are shown in **Table 2.3**.

Table 2.3: Loop-out clones created by Dr. Atlanta Cook

Name	Organism	Plasmid	Deletion
Nob1ΔL1	*S. cerevisiae*	pET28α	Δ151-159
Nob1ΔL2	*S. cerevisiae*	pET28α	Δ151-169
Nob1ΔL3	*S. cerevisiae*	pET28α	Δ151-176
Nob1ΔL4	*S. cerevisiae*	pET28α	Δ141-176
Nob1ΔL5	*S. cerevisiae*	pET28α	Δ141-159

1.2.2 20S rRNA transcription sequence cloning (helix 40 and D site)

Overlapping oligonucleotides were designed, ordered (IDT), and annealed prior to cloning to prevent self-annealing during further processing. The fragments were designed to prevent hairpin formation prior to cloning. The principle is shown in **Figure 2.1**.

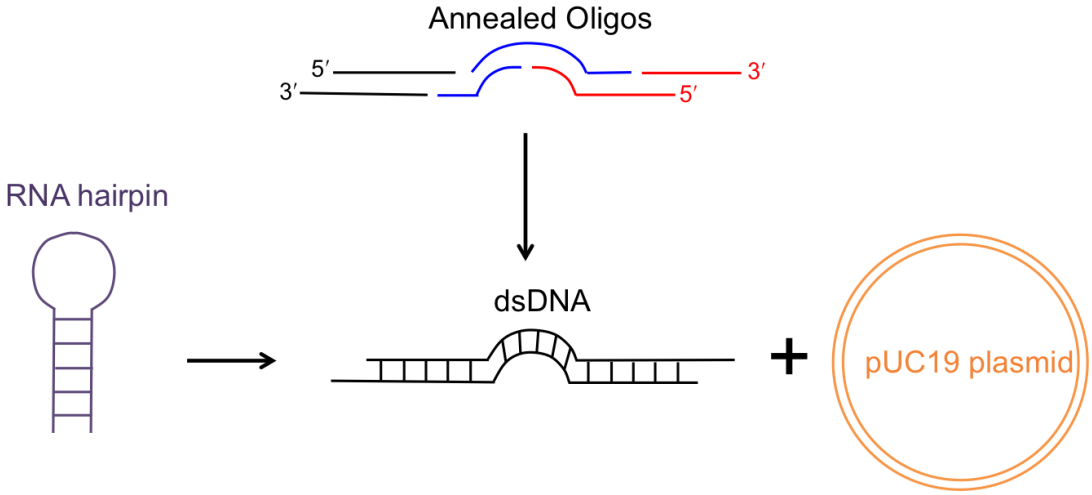


Figure 2.1: Cloning of transcription sequences for RNAs with secondary structures.
The sequence of the RNA hairpin (purple) has to be cloned as a double stranded DNA (dsDNA in black). However, the secondary structure of the RNA would also form in each strand of the DNA. Thus, initially, multiple overlapping oligonucleotides (shown at the top in black, blue, and red) are annealed with the corresponding oligonucleotides on the other strand (i.e. black-to-black, etc.). Then, the three overlapping fragments are annealed to each other to give the full-length dsDNA. This was then cloned into the pUC19 plasmid (orange).

Annealing reactions were performed in 10 µl containing 250 µM oligonucleotides and 1 x T4 polynucleotide kinase (PNK) buffer with ATP. These were heated to 90˚C for two minutes and placed directly on ice to remove secondary structures. 1 µl of T4 PNK (NEB) was added to each tube followed by incubation at 37˚C for 1 hour. The enzyme was then inactivated at 70˚C for 20 minutes in a water

bath. The samples were slowly cooled to room temperature in the water bath overnight. The annealed oligonucleotides were then cloned into the pUC19 plasmid.

The pUC19 plasmid was cut with EcoRI and BamHI (NEB) and gel purified using a Gel Extraction Kit (Qiagen). Equal volumes of the annealed oligonucleotides for each construct were pre-mixed and diluted five-fold. For the ligation reaction, 2 µl of plasmid were mixed with 2 µl of diluted mixed oligonucleotides, 1 µl (400 units) of T4 ligase and the T4 ligase buffer (NEB) in 10 µl total volume. The reactions were incubated at room temperature for 30 minutes and then, transformed into XL-1 *E. coli* strain using a standard transformation protocol for chemically competent cells (3 µl of each reaction were used). The transformed cells were selected using ampicillin resistance.

The pUC19 plasmids were extracted from 4 or 5 colonies for each construct using the MiniPrep kit (Qiagen). These were then digested with 2 units KpnI and ZraI (NEB) to test for loss of the KpnI site upon ligation with the insert (using 3 µl of extracted DNA). Positive clones were sent for Sanger sequencing and then re-transformed into XL-1 cells and extracted using a MaxiPrep kit (Qiagen) according to manufacturer's instructions.

1.3 Ligase-Independent Cloning (LIC)

The plasmids used can be seen in **Appendix II**. In order to prepare the plasmid, the GFP gene was excised using ZraI restriction enzyme (NEB). 2 µg of plasmid were incubated with 40 units of ZraI at 37˚C for 2 hours according to manufacturer's instructions. The linearized plasmid was then separated from the GFP fragment on 1% w/v agarose containing 1 x SYBR Safe DNA Stain (Invitrogen) for visualization in blue light. The plasmid was extracted using a Gel Extraction kit (Qiagen) according to the manufacturer's instructions.

The gene was amplified from the yeast genome using primers shown in **Appendix I**. The primers were designed to contain sequences that overlap with the target plasmid where it was cleaved with ZraI. The amplified fragment was separated on a gel and gel-purified in the same manner as the plasmid.

An overlap of 17 bases was created using T4 DNA polymerase (Novagen) which exhibits an exonuclease (3′ to 5′) activity. The nuclease is usually balanced by the polymerase activity (when a nucleotide is removed, the polymerase adds it back). However, this relies on all the nucleotides being available in solution. When only one of the nucleotides is present, others preceding the available one will be removed as the polymerase will not be able to add them back (Aslanidis & Dejong, 1990). This principle is used to create a precise overlap.

For the insert, 600 ng of DNA were incubated at room temperature for 30 minutes with 1 x T4 DNA polymerase buffer, 5 mM DTT, 2.5 mM dATP, and 1 unit of T4 DNA polymerase in a 20 µl volume. Thus, the overlap was created until the first adenine was reached (as dATP was available, this nucleotide was added back each time by the polymerase activity). The plasmid was processed in the same manner but using 450 ng of DNA, dTTP (instead of dATP to create complementary overlaps) and 1.5 units of enzyme. The final volume for these reactions was 30 µl. After 30 minutes, the enzyme was inactivated by incubating the reactions at 75˚C for 20 minutes.

The annealing reactions were performed by mixing 2 µl of the insert and 1 µl of the plasmid. This mix was incubated at room temperature for 10 minutes. Then, 1 µl of 25 mM EDTA was added and the mixture was incubated for a further 10 minutes at room temperature. Finally, the annealed plasmid was transformed (2 µl) into XL-1 cells using standard laboratory procedures. The overlap was longer than normally created with restriction digests. Thus, the insert and vector can stick together through the transformation. The breaks are then repaired by host ligases.

The positive clones were selected for on kanamycin or streptomycin-containing plates. Single colonies were then picked; the plasmids were extracted and tested for the insert using restriction digests. Two clones for each construct were sent for Sanger sequencing. **Table 2.4** lists all the constructs created this way. As a control for re-ligation without the insert, the same protocol was followed for the plasmid without adding the insert.

Dr. Uma Jayachandran also created the ScTsr1 constructs using this method.

Table 2.4: Constructs created using LIC

Insert	Organism	Plasmid
Dim2ΔN91	*S. cerevisiae*	pEC-KHC
Dim2ΔN51	*S. cerevisiae*	pEC-KHC
Dim2ΔN61	*S. cerevisiae*	pEC-KHC
Dim2ΔN99	*S. cerevisiae*	pEC-KHC
Dim2 FL	*S. cerevisiae*	pEC-SHT
Dim2 FL	*S. cerevisiae*	pEC-SGT
Dim2ΔN31	*H. sapiens*	pEC-KHC
Dim2ΔN42	*H. sapiens*	pEC-KHC
Nob1 FL	*H. sapiens*	pEC-KHC
Nob1 D10N	*H. sapiens*	pEC-KHC

1.4 Gibson Assembly Cloning

The gene inserts were amplified or (in the case of the tRNA/D site construct) ordered from NEB. The appropriate plasmid was amplified as well (primers shown in **Appendix I**). Standard PCR protocol was followed as for the other types of cloning. The inserts and plasmids were then separated on a 1% w/v agarose gel and extracted using a Gel Extraction kit (Qiagen).

The annealing reactions were performed in a 20 µl final volume. The amounts of DNA for both, the insert and the plasmid, were between 0.02 and 0.5 pmol with a 3-fold excess of the shorter fragment (i.e. the insert). This DNA mixture was incubated with 10 µl of the Gibson Assembly Master Mix (NEB). The reactions were made up to 20 µl using deionized water. Finally, the reactions were incubated at 50˚C for 1 hour.

The control reaction contained the same reagents apart from the insert, testing for any re-ligation of the plasmid. 2 µl of each of the reactions, including the control, were transformed into chemically competent XL-1 cells. Positive clones were selected using antibiotic resistance (ampicillin for pUC19 and kanamycin for pET28α). As with the other methods, several colonies were picked and grown up. The plasmids were extracted and tested using restriction digestion for the presence of the insert. Finally, two clones from each construct were sent for Sanger sequencing. Successful constructs are listed in **Table 2.5**.

Table 2.5: Constructs created using Gibson assembly.

Insert	Organism	Plasmid
tRNA/D site RNA sequence	*S. cerevisiae*	pUC19
Nob1 FL	*H. sapiens*	pET28α
Nob1 D10N	*H. sapiens*	pET28α

1.5 Expression Trials

The lysis buffer formulations used for the pull-downs of tagged proteins were as follows:

For Ni-NTA resin:

20 mM Tris-HCl pH8.0 at 4ºC

150 mM NaCl

0.5 mM β-mercaptoethanol

10 mM imidazole

For GSH resin:

20 mM Tris-HCl pH 7.5 at 4ºC

200 mM NaCl

1 mM DTT

After cloning, each of the constructs was subjected to small-scale expression trials. For this purpose, 25 ml of medium were inoculated with a swab of a plate containing cells transformed with the plasmid. Different strains were tested, typically four out of: BL21 (DE3), BL21 (DE3) pRIPL, BL21 (DE3) pLysS, T7 Express, and B834 (DE3). Initially, trials were performed in 2xTY and Superbroth media. However, 2xTY showed consistently better expression and was then used for further trials.

The inoculated culture was grown at 37˚C until an OD of 0.6-1.0 was reached. The cells were then cooled to 18˚C for one hour. A 1 ml fraction of the cells was then taken as the uninduced fraction. These cells were spun down at 13,000 rpm for 10 minutes. The supernatant was removed and 50 µl of sample buffer (100 mM Tris-HCl at pH 6.8, 10% v/v β-mercaptoethanol, 4% v/v SDS, 0.2% w/v bromophenol blue, 20% v/v glycerol) per 1.0 unit of OD were added to the pellet and

the cells were heated to 95˚C for 5 minutes. This fraction was stored at -20˚C until needed.

The rest of the culture was induced by addition of 0.2 mM IPTG and incubation at 18˚C overnight with shaking (250 rpm). The next day, the OD was measured and another 1 ml fraction was taken as the induced fraction. This was treated in the same way as the uninduced one. The rest of the culture was placed in a 50 ml Falcon tube and spun down at 4000 rpm for 10 minutes. The pellet was weighed and frozen at -20˚C until needed.

For the pull-down, an appropriate amount of lysis buffer was added to normalise for the cell mass (at least 2 ml) and the pellet was resuspended by vortexing. The cells were lysed by sonication using a Soniprep 150 sonicator (MSE) at 10 microns for 3 minutes (10 seconds on and 5 seconds off). 2 ml of the sonicated sample was centrifuged at 13,000 rpm for 10 minutes. The supernatant was placed in a new tube. 2 µl of the supernatant were added to 10 µl of sample buffer as the soluble fraction. 100 µl of sample buffer were added to the pellet making the insoluble fractions. Both of these were heated to 95˚C for 5 minutes and stored at -20˚C until needed.

The rest of the lysate was incubated for 1 hour at 4˚C (with shaking) with the appropriate resin for affinity binding of the protein of interest. Then, the beads were spun down at 7,000 rpm for 1 minute. They were washed 3 times with lysis buffer. The buffer was then removed and 30 µl of sample buffer was added to give the bead-bound fraction. It was heated to 95˚C and stored at -20˚C until needed.

All 5 fractions were loaded onto a denaturing polyacrylamide gel and separated by SDS-PAGE to show if the protein is expressed and if it is soluble. The gels were stained with Coomassie to visualise the proteins.

1.6 Fungal Protein Expression and Purification

1.6.1 ScNob1 – full-length

ScNob1 was expressed as a C-terminal hexahistidine fusion in the pET28α vector in the BL21 (DE3) *E. coli* strain (selection using kanamycin resistance). Later, it was also expressed in a pSL6 vector (also C-terminal hexahistidine fusion) in

StabyCodon SE1 (Eurogentec) *E. coli* strain (selection using ampicillin resistance) provided by the Tollervey lab. The cells were grown in 2xTY medium at 37˚C until OD of 0.6-1.0 was reached. They were then induced overnight with 0.2 mM IPTG at 18˚C. The culture was then harvested by centrifugation at 4500 rpm for 15 minutes. The pellet was flash frozen in liquid nitrogen until needed.

The cells harvested from 2 or 4 litres of culture were lysed in 50 ml of lysis buffer supplemented with protease inhibitors (1 ml of 100 mM AEBSF (Fluka) and an EDTA-free protease inhibitor tablet (Roche)) and 0.5 ml of 1mg/ml DNase (Sigma) in a Constant Systems Cell Disruptor (Constant Systems Limited) according to manufacturer's instructions. The lysis buffer consisted of 50 ml of 20 mM Tris-HCl (pH 8.0 at 4°C), 20 mM imidazole, 500 mM NaCl, 10% v/v glycerol, and 1 mM β-mercaptoethanol. The cleared lysate was collected by centrifugation at 50,000 g for 45 minutes and filtered (0.22 µm filter) prior to purification. All of the above steps were the same for all purification protocols although the buffers did change as described in the individual circumstances.

The cleared lysate was then loaded onto a His-Trap HP 5 ml column (GE Healthcare) on an AKTA purifier liquid chromatography system (GE Healthcare) pre-equilibrated with the lysis buffer. Absorbance at 280 and 260 nm was measured when possible. ScNob1 was eluted with a gradient of imidazole (up to 500 mM over 100 ml). 2 ml fractions were collected over the gradient. Fractions showing peaks in absorbance at 280 nm were examined using SDS-PAGE. Those that contained ScNob1 were pooled and loaded onto a HiPrep 26/10 Desalting column (GE-Healthcare) on the AKTA purifier. The protein was eluted into 20 mM Tris-HCl (pH 7.5 at 4°C), 100 mM NaCl, 10% v/v glycerol, and 1 mM DTT. Fractions displaying an absorbance peak at 280 nm were pooled. The same results were obtained when the desalting column was replaced with overnight dialysis into the same buffer.

Next, the protein sample was loaded onto a manually packed heparin sepharose column (GE Healthcare) on an AKTAprime plus system (GE Healthcare). It was eluted with a salt gradient (up to 1000 mM). Fractions showing a peak in absorbance at 280 nm were, again, examined using SDS-PAGE. ScNob1-containing fractions were pooled and concentrated to about 2 ml using a Vivaspin 20 concentrator (10 kDa cut-off) (Satorius Stedim Biotech).

Finally, the sample was loaded onto a pre-equilibrated (20 mM Tris-HCl pH 8.0 at 4°C, 150 mM NaCl, 5% v/v glycerol, and 1 mM DTT) HiPrep 16/60 S200 HR size exclusion chromatography (SEC) column (GE Healthcare) on the AKTA purifier system. Fractions encompassing the peak in absorbance at 280 nm were examined using SDS-PAGE and pooled. The pure ScNob1 protein was, again, concentrated using the Vivaspin 20 concentrator to give a final concentration of about 8-10 mg/ml protein. The protein was supplemented with 30% v/v glycerol, flash-frozen in liquid nitrogen, and stored at -80˚C.

The protocol was adapted to include a cation exchange step: a Mono S 5/50 GL column (GE Healthcare) or a manually packed SP Sepharose (GE Healthcare) column with a higher binding capacity, instead of the heparin step. The HiPrep SEC column was replaced with a HiLoad Superdex 200 16/60 column (GE Healthcare). In all the following purification protocols, the fractions showing peaks on the chromatograms were examined with SDS-PAGE after every purification step.

1.6.2 ScNob1 D15N – Catalytically inactive mutant

Expression was performed using a construct obtained from the Tollervey lab in the same manner as for the wild-type protein. Initially, the lysis was also performed in the same manner and in the same buffer. Later, the Tris-HCl concentration was raised to 40 mM and a 5-minute sonication step (cycles of 10 seconds on followed by 5 seconds off) was added after cell disruption to remove aggregates. This was performed in a Soniprep 150 sonicator (MSE) at 10 microns.

The first step of purification was performed in-batch on Nickel resin (GE Healthcare). For this purpose, the beads were incubated with the cleared lysate on a spinning wheel for 1 hour at 4˚C. An XK 16/20 column (GE Healthcare) was then manually packed and connected to an AKTAprime (GE Healthcare) system. The protein was eluted in the same manner as the other construct on the pre-packed column.

The mutant ScNob1 was then dialysed overnight into 20 mM HEPES pH 7.5, 150 mM NaCl, 5% v/v glycerol, and 1 mM β-mercaptoethanol. DTT was not used in the buffers as it can chelate metals, such as zinc, which might be required for RNA binding via the zinc ribbon domain of Nob1 (Krezel et al, 2001). The sample was

injected onto a pre-equilibrated, Resource S column (GE Healthcare). This was later changed to a manually packed SP Sepharose (GE Healthcare) column with a higher binding capacity. The flowthrough was washed out and ScNob1 D15N was eluted in a 100 ml gradient of dialysis buffer supplemented with 1000 mM NaCl. The fractions showing a peak on the 280 nm absorbance spectrum were analysed by SDS-PAGE. Fractions containing the protein of interest were pooled and concentrated to a volume of 2 ml.

Next, the sample was injected onto a pre-equilibrated (20 mM HEPES pH 7.5, 150 mM KCl, 5% v/v glycerol, and 1 mM β-mercaptoethanol) HiLoad Superdex 200 16/60 column (GE Healthcare) column. Again, the fractions obtained were examined by SDS-PAGE. The protein was pooled and concentrated to about 2 mg/ml. Later, the wild-type protein was purified in the same manner to allow easier comparisons.

1.6.3 ScNob1 – loop-out mutants

Initially, ScNob1ΔL3 and ΔL4 were purified. The same protocol as for the D15N mutant was used with the exception that the Resource S column was replaced with a Resource Q anion exchange column (GE Healthcare). This change was necessary as the removed portion of the protein was highly charged and its deletion significantly changed the pI of the constructs. All the buffers were the same.

Then, with the help of a summer intern (Michal Merdas), the other three mutants were also purified using the same protocol.

1.6.4 ScDim2 – full-length and different constructs

Purification of several different constructs was performed. **Table 2.6** summarises how far each of the protocols was taken and what methods were used.

For each of the protocols, the cells were lysed as previously described for ScNob1. The lysis buffer consisted of the respective buffer A for the first step of purification supplemented with protease inhibitors and DNase as with ScNob1. All initial affinity steps were performed in-batch and the columns were packed manually prior to connecting them to the AKTAprime system (GE Healthcare). In order to

achieve binding, the beads were washed with lysis buffer and then incubated with the cleared lysate for at least 1 hour at 4˚C.

Table 2.6: Summary of the purification methods used on different ScDim2 constructs.

Construct	Purification step 1	Purification step 2	Purification step 3
Dim2ΔN91	GSH resin	-	-
Dim2 FL	GSH resin	Mono S / Mono Q	-
Dim2 FL and ScNob1 co-lysis	GSH resin	-	-
Dim2 FL	Ni^{2+} NTA	-	-
Dim2ΔN51	Ni^{2+} NTA	Mono S	SEC

For the purification of Dim2ΔN51 see HsDim2 protocol later. All other constructs were purified as follows. For GSH affinity purification, the buffer (A) used consisted of 20 mM Tris-HCl (pH 7.5 at 4˚C), 500 mM NaCl, and 1 mM DTT and the protein was eluted with 20 mM reduced GSH. For the purification of full-length His-tagged protein, buffer A consisted of 20 mM Tris-HCl (pH 7.5 at 4˚C), 500 mM NaCl, 20 mM imidazole, and 1 mM β-mercaptoethanol. The bound protein was eluted with a gradient of 20-500 mM imidazole over 100 ml.

After the affinity step, the respective tags were cleaved overnight using 0.5 mg of 3C protease during dialysis into low salt buffer. This buffer consisted of 20 mM Tris-HCl (pH 7.5 at 4˚C), 50 mM NaCl, and 1 mM DTT.

The full-length protein purification was then extended into an ion exchange step. For the GST-tagged version, a Mono S column was used initially. However, a very small amount of the protein bound. Therefore, the Mono Q anion exchange column was also used. The protein was eluted with a 100 ml gradient going up to 1000 mM NaCl. The ScDim2 protein had a lot of RNA contamination so this was not continued.

The His-tagged protein purified was subjected to nuclease digestion to remove the RNA. The dialysed protein was supplemented with $MnSO_4$ to a final concentration of 6 mM and 10 units of Cyanase nuclease (RiboSolutions, Inc.). 20 µl fractions were taken every 30 minutes for 3 hours. The RNA was digested but the protein precipitated. Therefore, the purification was discontinued.

1.6.5 CtNob1 – full-length

The expression pRSET-A plasmid containing the CtNob1 gene (**Appendix III**) was ordered as a codon optimized construct from GeneArt. The protein was expressed in chemically- or electro-competent BL21 (DE3) cells in 2xTY medium. The cells were grown to OD 0.6-1.0 at 37˚C and then cooled to 18˚C for one hour. Induction was carried out with 0.2 mM IPTG overnight. 4 litres of culture were used for purification.

The initial protocol was identical to the ScNob1 protocol. Over time this was developed to include an anion exchange step (Mono Q 5/50 GL column, GE Healthcare). The dialysis buffer contained 50 mM salt but was otherwise the same. The HiLoad Superdex 200 16/60 column (GE Healthcare) was used for size exclusion chromatography.

1.6.6 CtDim2 – full-length

As with the CtNob1 plasmid, the construct was ordered from GeneArt (**Appendix III**). Expression trials were performed; however, little or no expression was seen in BL21 (DE3), BL21 (DE3) pRIPL, T7 Express, and B834 (DE3) cell strains.

1.6.7 ScTsr1 – loop-out constructs

Dr. Uma Jayachandran performed the initial expression and purification of the constructs. Once crystals were obtained, the established protocols were used to make seleno-methionine labelled protein. For this purpose, the expression plasmid containing ScTsr1ΔN46Δ410-476 was transformed into the B834 (DE3) expression cell line. The cells were plated and incubated at 37˚C overnight to obtain colonies.

The next morning, the colonies were picked and placed in growth media supplemented with seleno-methionine. The medium consisted of 2 x M9 medium (37.4 mM NH_4Cl, 44.1 mM KH_2PO_4, and 95.8 mM anhydrous Na_2HPO_4) supplemented with 0.4% w/v glucose, 40 µg/ml amino acids (with seleno-L-methionine instead of methionine), and 1 µg/ml vitamins (riboflavin, niacinamide, pyridoxine monohydrochloride, and thiamine). Two litres were grown up to OD 0.6-

1.0 at 37˚C, cooled to 18˚C, and induced overnight with 0.2 mM IPTG. The pellets were then spun down and kept at -80˚C as described for other proteins.

Cell disruption was performed as described previously. The initial Ni-NTA affinity purification step was performed in-batch as described for ScNob1 D15N purification. Fractions containing the protein were pooled, mixed with 0.5 mg 3C protease, and dialysed into 20 mM Tris pH 8.0 at 4˚C, 600 mM NaCl, 5% v/v glycerol, and 1 mM DTT overnight. The protein was then concentrated to 2 ml and injected onto a HiLoad Superdex 16/60 S200 column (GE Healthcare) pre-equilibrated in 20 mM Tris pH 8.0 at 4˚C, 800 mM NaCl, 5% v/v glycerol, and 1 mM DTT.

Fractions containing the protein were separated on 15% SDS-PAGE and analysed. The fractions not overlapping with 3C protease were pooled and concentrated to 10 mg/ml. The protein was divided into 25 µl fractions, flash-frozen in liquid nitrogen, and stored at -80˚C until needed.

1.6.8 ScFap7 and ScDim1

ScFap7 was purified by Dr. Uma Jayachandran while Dim1 was purified by Anees Basha. They were stored in 30% v/v glycerol at -80˚C and -20˚C respectively until required.

1.7 Human Protein Expression and Purification

The sequences for the human proteins were obtained from the laboratory of Dr. Nick Watkins at the University of Newcastle. The constructs obtained in this manner are: HsDim2-FL in pGEX-6P-1 plasmid (GST-tagged) and HsNob1-FL in pET100 (His-tagged). These were supplied along with expression protocols and a one-step purification protocol used routinely in Dr. Watkins' laboratory. Details of the purification methods are listed below. The HsDim2 construct used for expression and purification was not altered. HsNob1 was re-cloned into different vectors for easier tag cleavage or to eliminate degradation products. The clones are listed earlier in the methods section.

1.7.1 HsDim2 – full-length

The protocol provided by the Watkins lab established the first step of purification. BL21 (DE3) cells were grown, induced, harvested, and lysed as described for the yeast Nob1. The lysis buffer consisted of 20 mM Tris-HCl (pH 8.0 at 4°C), 300 mM KCl, 10% v/v glycerol, 0.1% v/v Tween20, and 1 mM DTT (GSH Buffer A). As previously, it was supplemented with the protease inhibitors and DNase. The cleared lysate was bound in batch to GSH resin (GE Healthcare) at 4˚C with mixing for two hours. The resin was then spun down at 500 x g for 5 minutes. The supernatant was removed and the resin was washed three times with 50 ml of lysis buffer (each time it was spun down and the supernatant was removed).

Then, the GST-tagged HsDim2 was eluted using the above buffer supplemented with 50 mM reduced glutathione as recommended by the Watkins lab. 10 ml of this buffer were added to the resin. The mixture was incubated at 4˚C for 10 minutes with shaking. Then, the sample was spun down as before and the supernatant removed. This was repeated four more times to give five elution fractions. These samples were separated on a 12.5% SDS-PAGE gel. Fractions containing HsDim2 were pooled and dialysed overnight into Mono S Buffer A (20 mM HEPES (pH 7.5), 100 mM KCl, 10% v/v glycerol, and 1 mM DTT) in the presence of 0.5 mg 3C protease.

The dialysed protein was loaded onto a Mono S 5/50 GL column (GE Healthcare) and eluted with a gradient of Mono S Buffer B (i.e. Mono S Buffer A containing 1000 mM KCl). The gradient lasted 100 ml. The fractions containing the resulting peak and surrounding regions were examined on a 12.5% SDS-PAGE gel. The appropriate fractions were pooled and concentrated to less than 2.5 ml using the Vivaspin 20 (10 kDa cut-off) (Satorius stedim biotech) concentrator for injection onto a HiLoad Superdex 200 16/60 SEC column (GE Healthcare).

The SEC was carried out in the following buffer: 20 mM Tris-HCl (pH 7.5 at 4˚C), 150 mM KCl, 5 % v/v glycerol, and 1 mM DTT. A single peak was obtained and examined on a 12.5% denaturing gel. The fractions containing HsDim2 were pooled and concentrated to give a final concentration of 8.0-10.0 mg/ml. Protein to be used within a week was stored at 4˚C while long-term storage was carried out at -80˚C in 30% v/v glycerol.

His-tagged ScDim2ΔN51 was then purified in the same manner with the exception that nickel resin in-batch purification (as described earlier for ScNob1 D15N) was used as the first step. The lysis buffer consisted of 20 mM Tris-HCl (pH 8.0 at 4°C), 300 mM KCl, 20 mM imidazole, 10% v/v glycerol, 0.1% v/v Tween20, and 1 mM β-mercaptoethanol. The protein was eluted with a 100 ml gradient of imidazole reaching 500 mM. The following steps and buffers were the same as for the human protein.

1.7.2 HsDim2 – proteolysed fragment

A truncated version of HsDim2 was purified following preparative proteolysis using GluC (Sigma) as the protease. The expression, lysis, and first two steps (GSH affinity and ion exchange) of purification were performed as for the full-length protein. Then, the protein was concentrated to 1.25 ml (3.8 mg) and mixed with GluC (0.005 mg GluC per 1 mg of HsDim2) and 400 µl of 20 mM HEPES pH 7.5, 50 mM NaCl, and 10 mM $MgSO_4$. This was incubated overnight at 4˚C. These conditions were established following small-scale limited proteolysis (see standard laboratory protocols). The next morning, 110 µl of 100 mM AEBSF protease inhibitor (Fluka) was added to the reactions. The sample was immediately injected onto a SEC column as for the full-length protein.

Two major species were obtained seen as a double peak on the SEC profile. These were concentrated separately and used for crystallisation screens.

1.7.3 HsNob1 – full-length

The human Nob1 was re-cloned into a pEC-KHC plasmid as described previously, to allow for easier tag cleavage with the 3C protease. The expression was performed in 2xTY medium in BL21 (DE3) cells. The pelleted cells were lysed as for the yeast protein. The buffers for the lysis and the first step of purification (nickel affinity) were the same as for ScNob1. The protein was bound to the resin in-batch. The cleared lysate was incubated with 20 ml nickel affinity resin (GE Healthcare) at 4˚C for at least 1 hour with mixing. The beads were then spun down and manually packed into a XK 16/20 column (GE Healthcare). The eluted HsNob1 was dialysed

into 20 mM Tris-HCl (pH 8.0 at 4˚C), 100 mM NaCl, and 1 mM DTT overnight in the presence of 0.5 mg of 3C protease to allow tag cleavage.

The protein was then subjected to Mono Q anion exchange purification where it was eluted off the column with a 100 ml gradient reaching 1000 mM NaCl. The protein was found to bind to the column but further purification with this construct was not performed due to the presence of degradation products.

To prevent the binding of degraded protein to the nickel column, the human Nob1 gene sequence was cloned into a pET28α vector to contain a non-cleavable C-terminal His-tag. The plasmid was transformed into BL21 (DE3) cells and grown up, harvested, and lysed in the same manner as previous clones used for purification. The binding to the nickel resin and the packing of the column were performed the same as for the above construct.

The column was attached to an AKTAprime system. The unbound protein was washed out and the column was washed with 50 ml lysis buffer supplemented with 2 mM ATP and 10 mM $MgSO_4$ to remove any chaperone proteins. The bound HsNob1 was then eluted using a 100 ml imidazole gradient going up to 500 mM, as for the yeast protein. The column was run at 2 ml/min and 2-5 ml fractions were collected. The fractions containing the peaks at 280 nm were separated on a 12.5% denaturing SDS/acrylamide gel and fractions containing the HsNob1 were pooled and dialysed overnight into 20 mM Tris-HCl at pH 8.0 (at 4˚C), 100 mM NaCl, 10% v/v glycerol, and 1 mM DTT.

Next the protein was injected onto a pre-equilibrated Mono Q 5/50 GL column (GE Healthcare) and eluted as described for the previous proteins. Finally, the HsNob1 was concentrated to less than 250 µl and injected onto a HiLoad Superdex 200 10/30 column (GE Healthcare) pre-equilibrated in 20 mM Tris-HCl pH 7.5 (at 4˚C), 150 mM NaCl, 5% v/v glycerol, and 1 mM DTT. The rest of the treatment was the same as for the other proteins (i.e. concentrating and freezing).

1.8 rRNA *In Vitro* Transcription

In vitro transcription was either performed using the MEGAshortscript kit (Ambion) or the T7 polymerase purified in the lab. The details for each of the RNAs can be found in the following sections.

1.8.1 Expression and Purification of T7 polymerase

The His-tagged T7 Polymerase has been previously cloned into an M3 plasmid and was expressed in Superbroth medium (supplemented with Kanamycin and Ampicillin) in M15 cells as described for ScNob1. It was purified using a two-step protocol: cobalt affinity column followed by a hydroxyapatite (ion exchange) step, as previously established (Arts et al, 1997).

Cell lysis was performed in the same way as for the other proteins described. The lysis buffer consisted of 20 mM Tris-HCl (pH 8.0 at 4˚C), 500 mM NaCl, 5 mM imidazole, and 1 mM β-mercaptoethanol supplemented with DNase and protease inhibitors as for the other proteins. The cobalt affinity purification was performed in batch using cobalt affinity resin (Sigma). An XK 16/20 column (GE Healthcare) was manually packed and attached to the AKTAprime system as described previously. The protein was eluted with a 100 ml gradient of imidazole (going up to 500 mM). The fractions containing the T7 polymerase were pooled and dialysed into hydroxyapatite buffer A (20 mM potassium phosphate buffer at pH 7.5, 50 mM KCl, 0.3 mM $CaCl_2$, and 3 mM β-mercaptoethanol).

The dialysed protein was filtered to remove any precipitate and injected onto a manually packed, pre-equilibrated hydroxyapatite column. The protein was eluted with a 100 ml gradient up to 7.5% w/v ammonium sulphate. To remove the salts, the T7 polymerase was dialysed again into 20 mM potassium phosphate buffer at pH 7.5, 150 mM NaCl, and 1 mM DTT. It was then concentrated to 10 mg/ml. A final concentration of 50% v/v glycerol was added (to give a T7 polymerase concentration of 5 mg/ml) and 1 ml aliquots were frozen in liquid nitrogen and stored at -80˚C until needed.

The activity of the polymerase was tested on a small scale using the same protocol as was given by the manufacturer for the MEGAshortscript kit (Ambion).

However, all reagents were made in the lab, including the nucleotide solutions and the 10 x transcription buffer (400 mM Tris-HCl pH 8.0, 500 mM $MgCl_2$, 10 mM spermidine, 0.1% v/v Triton, and 50 mM DTT).

1.8.2 Helix 40 – Small Scale

The ScNob1 interaction with the 20S rRNA has been previously mapped, using crosslinking (Granneman et al, 2010), to a single loop (helix 40). To investigate this, three sequences encompassing this loop were transcribed *in vitro* (shown in chapter 4). For this purpose, the sequences were cloned into the pUC19 plasmid as described earlier.

For *in vitro* transcription, 500 µl of each of the three plasmids were linearized using 100 units of SmaI (NEB) overnight in an 800 µl reaction. Linearization was tested on a 1% w/v agarose gel where the bands were visualized with SYBR Safe Stain (Invitrogen). The linear plasmids were extracted using a standard phenol extraction protocol.

Test *in vitro* transcriptions were performed using the MEGAshortscript kit (Ambion) in 20 µl reactions using 25 nM of the templates according to manufacturer's instructions. This was then scaled up to 80 µl reactions. 3 µl of the reaction contents were examined on a 12% denaturing urea/acrylamide gel (20 x 20 cm, 0.75 cm thick acrylamide, 1 x TBE, 7 M Urea), which was run at 20W for 30 minutes. The gel was stained in a 1 x SYBR Safe Stain (Invitrogen) solution for 5 minutes for visualization. The rest of the RNA was extracted using a standard phenol extraction protocol.

1.8.3 D site – Small scale

Several D site sequences were designed, as shown in chapter 4. Some were transcribed from oligonucleotides while others from plasmids (see **Appendix IV**). For transcription from oligonucleotides, the two complementary strands were ordered from IDT and annealed overnight. This was achieved by mixing equimolar amounts and placing them for 10 minutes in a beaker containing water heated to 95˚C. The heat was then switched off and the samples and beaker were allowed to

cool to room temperature overnight. For transcription from a pUC19 plasmid, linearization of the DNA was performed as described for Helix 40 RNAs.

In vitro transcription was performed using T7 polymerase purified in the lab. Each transcription reaction contained 0.5-5 µM template, 25 mM nucleotides, and 5 µg of polymerase in 20 µl of buffer (40 mM Tris-HCl pH 8.0, 50 mM $MgCl_2$, 1 mM spermidine, 0.01% v/v Triton, and 5 mM DTT). The reactions were mixed at room temperature and incubated at 37˚C for 3 hours. Often, the concentration of the template had to be optimised.

To end the reaction, 1 µl of TURBO DNase (Ambion) was added and the tubes were incubated at 37˚C for a further 1 hour. The RNA was then analysed by denaturing urea/acrylamide gel electrophoresis (acrylamide, 1 x TBE, 7 M Urea) on a 12-16% gel. The 20 x 20 cm gel was run at 20 W in 1 x TBE for 45 minutes.

The transcription was scaled up accordingly. When required, the RNA was gel purified as described for the radioactive labelling protocol in the EMSA section of the methods.

1.8.4 D site – Large scale and Purification

For large-scale transcription, the above protocol was proportionally scaled up. For the yeast D site with the short hairpin, 1.25 µM template was used. The rest of the reagents were the same, but in a 5 ml volume in a 50 ml Falcon tube. Again, the reaction was incubated at 37˚C for 3 hours. The tube was then spun down at 4,000 rpm at room temperature to remove the precipitated inorganic phosphate. The supernatant was filtered with a 0.22 µm filter prior to injection onto a pre-equilibrated (20 mM HEPES pH 7.5, 150 mM KCl, 5% v/v glycerol, 5 mM $MgCl_2$, and 1 mM DTT) HiLoad Superdex S200 16/60 SEC column. The column was run at 1 ml/min for 1.5 column volumes and 1 ml fractions were collected. This removed the spermidine from the sample.

The fractions were analysed on a 16% denaturing gel and the ones containing the right fragment were pooled. Further purification was achieved using anion exchange (Mini Q column, GE Healthcare). The column was pre-equilibrated with 20 mM HEPES pH 7.5, 150 mM KCl, and 5 mM $MgCl_2$. A small injection of 100 µl of RNA was used to establish the best elution conditions (column was run at 0.5

ml/min, and 0.25 ml fractions were collected). Initially, the RNA was eluted with a 20 column volume linear gradient going up to 1000 mM KCl. Absorbance at 260 and 280 nm was measured during the runs.

The optimised elution protocol consisted of binding at 150 mM KCl followed by a wash at 350 mM KCl for 3 column volumes. Then, there was a 4-column volume gradient from 350 to 450 mM KCl. A wash at 450 mM salt was carried out for 3 column volumes. Finally, a gradient from 450 to 700 mM KCl was used to elute the last of the RNA over 4 column volumes. To clean the column, a final 1 column volume of 1000 mM salt was used. The different peaks were examined on a 16% denaturing gel and pooled separately. Peak 3 was then used for purification of the complex (see later).

1.9 Primer Extension to Determine rRNA Secondary Structure

In order to determine the fold of the helix 40 *in vitro* transcribed sequences, longer versions were created to enable primer binding. The same protocol as described earlier was followed except that the plasmids were linearized with HindIII to add an extension. A primer was designed to bind to this extension on the created RNA (AGCTTGCATGCCTGC). The primer was labelled with radioactive (^{32}P) γ-ATP. For this purpose, 2.5 µM primer was incubated at 37˚C for 1 hour in the presence of 20 units PNK (NEB), 1 x PNK buffer, 10 mM DTT, and 10 uCi ^{32}P-ATP in a final volume of 20 µl. Then, 30 µl of sterile water were added and the sample was spun through a ProbeQuant G-50 Mini column (GE Healthcare).

The RNA was diluted (2 µg in 100 µl) in 50 mM Tris-HCl pH8.0, 150 mM NaCl, 1.5 mM $MgCl_2$. It was refolded by heating to 65˚C for 10 minutes in a heating block and then cooling to room temperature. 15 mM 1-methyl-7-nitroisatoic anhydride (1M7) was added to each of the RNA samples and mixed. 1M7 modifies the unpaired bases. The nucleic acids were then extracted with a standard phenol/chloroform extraction followed by a chloroform extraction and ethanol precipitation. The extraction was also performed for control samples, which were set up in the same manner but were not refolded or treated with 1M7. These would show natural transcription stop points.

The primer (1 pmol) was mixed with the extracted RNA (15 ng) in a 2.75 µl volume. This was heated to 85˚C for 3 minutes. Then, 0.5 mM dNTPs, 1 x reverse transcriptase buffer, 5 mM DTT, 10 units of RNasin, and 50 units Superscript III Reverse Transcriptase (Invitrogen) were added to make up the reaction volume to 5 µl. This was incubated at 50˚C for 50 minutes to allow primer extension. The reactions were stopped by addition of 5 µl of formaldehyde loading dye. The same protocol was followed for the control reactions. The transcription should stop more frequently before bases modified with 1M7.

Sequencing ladders for each transcript were also prepared. 5 µg of plasmid DNA were mixed with 0.2 M NaOH and 0.2 mM EDTA. This was incubated at 37˚C for 30 minutes. The plasmids were then ethanol precipitated and resuspended in 35 µl of sterile water. 7 µl of each plasmid were mixed with 1 pmol of the radiolabelled primer and 1 x Sequenase buffer. The reactions were incubated at 65˚C for 10 minutes and allowed to cool to 10˚C in a heating block. 7 mM DTT, 0.4 µl of labelling mix, and 26 units Sequenase (Affymetrix) were added to give a final volume of 15 µl. Finally, 3.5 µl of these reactions were added to 2.5 µl of each of 4 termination mixtures that were pre-heated to 37˚C. This was incubated at 37˚C for 10 minutes and stopped by adding 4 µl of stop solution, as suggested by the manufacturer.

The primer extension reactions and corresponding ladders were separated on a 6% denaturing gel. They were first heated to 95˚C for 2 minutes and then chilled on ice. The gel was run at 30W for about 1.5-2 hours. The gel was dried and exposed to a phosphor-screen overnight. The screen was then scanned using Fuji BAS-5000 phosphoimager.

An attempt was made to see if ScNob1 could protect the RNA from modification by binding to the unpaired regions. For this purpose, after the refolding step, the RNA was incubated with ScNob1 for 30 minutes. The rest of the protocol remained the same.

1.10 Complex Formation and Purification

1.10.1 Semi-analytical sixe exclusion chromatography (SEC)

Initially, CtNob1 was used with helix 40 RNA (from *C. thermophilum* or *S. cerevisiae*). 100 µg (1.90 nmol) of protein were mixed with either 1.2 x or 4 x molar excess of the corresponding RNA in a 110 µl volume in 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 4 mM $MgCl_2$, and 1 mM β-mercaptoethanol.

The samples were incubated at room temperature for 1 hour. Then, 110 µl were injected into a 100-µl loop attached to an AKTA purifier system with a pre-equilibrated Superdex 200 GL 10/300 column (GE Healthcare). The protein and RNA were eluted with 1.2 column volumes of the buffer run at 0.5 ml/min (0.5 ml fractions were collected). Fractions of interest were separated on gradient SDS-PAGE (Bio-Rad) and silver-strained (see later for protocol).

The rest of the semi-analytical SEC experiments described were performed in the same manner. However, the amounts of protein used and final volumes changed (summarised in **Table 2.7**). The buffer for the other experiments consisted of 20 mM HEPES pH 7.5, 150 mM KCl, 5% v/v glycerol, 5 mM $MgCl_2$, and 1 mM β-mercaptoethanol.

Table 2.7: Different analytical SEC experiments performed after the first one described above. The RNAs used were either at 1.2 or 4 times molar excess relative to the proteins. The relevant amounts are shown later with the results. All proteins used were in a one-to-one molar ratio with respect to each other.

Experiment	Amount of protein (µg)	Amount of protein (nmol)	Final volume to be injected (µl)	Volume of loop (µl)
ScNob1 and HsDim2	100 (HsDim2) 186 (ScNob1)	3.53	60	50
ScNob1 and helix 40	186	3.53	60	50
ScNob1 and D site RNAs	50	0.95	60	50
HsDim2 and D site RNA	50	1.76	60	50
ScNob1, HsDim2 and D site RNA	50 (HsDim2) 93 (ScNob1)	1.76	60	50
ScDim2ΔN51 and D site RNAs	100	3.65	120	100
ScNob1 and ScDim2ΔN51	100 (ScDim2) 192 (ScNob1)	3.65	120	100
ScNob1 and ScDim1	192 (ScNob1) 131 (ScDim1)	3.65	240	200
ScDim2ΔN51 and ScDim1	100 (ScDim2) 131 (ScDim1)	3.65	240	200
ScNob1 and ScFap7	114 (ScNob1) 50 (ScFap7)	2.10	60	50

1.10.2 ScNob1 and D site RNA complex purification

In order to purify the complex, the protein was obtained separately first. Thus, the affinity step and the ion exchange were performed as described for ScNob1. Once pure protein was obtained, it was mixed with 3 times molar excess of Peak 3 (see earlier) D site RNA. A 5 x buffer solution was added to the protein first to give a final buffer composition of 20 mM HEPES pH 7.5, 150 mM KCl, 5% v/v glycerol, 5 mM $MgCl_2$, and 1 mM β-mercaptoethanol in a 2 ml final volume.

The sample was incubated at room temperature for 1 hour. Then, it was injected onto a pre-equilibrated HiLoad Superdex S200 16/60 column (GE Healthcare). It was run at 1 ml/min for 1.5 column volumes while 1 ml fractions were collected. Absorbance at 280 and 260 nm was measured and standard SDS-PAGE was used to analyse the protein content of the peak obtained on the chromatogram. The chromatogram was also aligned to that of ScNob1 alone to compare the elution volumes of the peaks. The part of the peak that did not overlap with the protein alone chromatogram was pooled and concentrated to 8.2 mg/ml.

1.11 Biochemical Assays

1.11.1 Thermal Denaturation

Different proteins were subjected to thermal denaturation in different buffers in order to examine changes in protein stability. The concentrations used were initially optimized based on the best signal-to-noise ratio while using the least protein. The concentrations used for the different proteins and the variables tested for that sample can be seen in **Table 2.8**. The assay was performed in plastic, clear 96-well iQ Real-Time PCR plates (Bio-Rad). Each well contained 5 x Sypro Orange (Sigma Aldrich), testing buffer, and the necessary amount of protein in a final volume of 50 µl. The proteins were then subjected to incremental increases in temperature (0.5˚C every 30 seconds from 20˚C to 80˚C). The change in fluorescence was monitored using an iQ5 rtPCR Thermocycler (Bio-Rad). The data was then analysed using the iQ5 Optical System Software (BioRad) and re-plotted in Microsoft Excel.

Table 2.8: Variables tested in thermal denaturation for each protein and the concentration at which the protein was used.

Protein	Concentration (µM)	Variables tested
ScNob1	2	pH (8.0-6.5), glycerol (0-10% v/v), NaCl (50-250 mM), zinc acetate (0-10 µM)
CtNob1	2	pH (8.0-6.5), glycerol (0-10% v/v), NaCl (50-250 mM), zinc acetate (0-10 µM)
HsDim2	3	pH (8.0-6.5), glycerol (0-10% v/v), KCl (50-250 mM), $MgCl_2$ (0-20 mM)

1.11.2 Electrophoretic mobility shift assays (EMSAs)

1.11.2.1 5′ labelling

300 pmol of each of the RNAs were gel-purified according to standard laboratory procedures. The RNA was then resuspended in 15 µl of sterile water. After gel purification, the labelling reactions were assembled. 100 pmol of RNA were mixed with 20 units of polynucleotide kinase (PNK, NEB), 1 x PNK buffer, and 10 uCi of ATP (γ-^{32}P) in a final volume of 20 µl. These reactions were incubated at 37˚C for 2.5 hours. Then, 180 µl of splicing dilution buffer (100 mM Tris-HCl at pH 7.5, 150 mM NaCl, 300 mM NaOAc at pH 5.2, 10 mM EDTA, and 1% w/v SDS) were added and the RNA was extracted using a standard phenol/chloroform extraction and ethanol precipitation procedure. The samples were dried and resuspended in 50 µl of sterile water (2 pmol/µl final concentration).

1.11.2.2 Helix 40 RNA and ScNob1 or CtNob1

Initially, a standard TBE buffer system was used. 2 pmol of RNA were mixed with varying amounts of protein (0-400 pmol) in a final volume of 10 µl (in a buffer consisting of 20 mM HEPES pH 7.5, 150 mM NaCl, 4 mM $MgCl_2$, and 1 mM β-mercaptoethanol). These reactions were then incubated at 4˚C, room temperature, or 37˚C for 30 minutes. 2 µl of non-denaturing dye (0.25% w/v bromophenol blue, 0.25% w/v xylene cyanole FF, and 50% v/v glycerol) were added and 4 µl were loaded onto an 8% native acrylamide gel (0.5 x TBE, 8% acrylamide). The gel was run at 200V at 4˚C or at room temperature until the dye was halfway down the gel.

Once finished, the gel was wrapped in cling film and taped inside a cassette under a phosphoscreen for exposure. The cassette was stored at -80˚C overnight. The next day, the gel was removed and the phosphoscreen was scanned using a Storm 860 Molecular Imager (GE Healthcare).

1.11.2.3 D site and ScNob1

RNA fragments based on the ones found to interact with ScNob1 in the semi-analytical SEC were obtained from IDT (sequences can be seen in **Table 4.6** in the results section). Initially, the same EMSA protocol was used as for helix 40. However, since the complex could not enter the gel, the percentage of acrylamide was reduced to 4%. Also, different running buffers were tested (McLellan, 1982). **Table 2.9** lists the different buffers used. Once finished running, instead of freezing the gel, it was dried according to the manufacturer's instructions in a GD 2000 Slab Gel drier (Hoefer). The phosphoscreen was exposed overnight and then scanned using a Typhoon FLA 7000 Imager (GE Healthcare). An EMSA using the ScNob1 D15N mutant protein was performed in the same manner. All the EMSAs were repeated with different batches of purified protein.

Table 2.9: The different buffers used for EMSAs.
The standard names of the buffers and final concentrations of the components are shown. CAPS is also known as 3-(Cyclohexylamino)-1-propanesulfonic acid.

Buffer name	pH	Basic component	Acidic component	Additional component
0.5 x TBE	8.3	45 mM Tris base	45 mM Borate	1 mM EDTA
1.0 x TC	9.4	60 mM Tris base	40 mM CAPS	-
1.0 x AC	10.2	37 mM Ammonia	20 mM CAPS	-

1.11.3 Cleavage assays

Initially, a cleavage assay was set up using 2 pmol of DS5 RNA and 10 pmol of ScNob1 in 10 µl. The buffer was the same as for the EMSA with an addition of 5 mM $MnCl_2$ as was shown to be required by the Tollervey lab (Lebaron et al, 2012). The samples were incubated at room temperature for 2 hours with fractions being taken every 15 minutes. To stop the reaction, 10 µl of denaturing dye (90% v/v formamide, 10 mM Tris-HCl pH 8.0, 2 mM EDTA, 0.05% w/v bromophenol blue, and 0.05% w/v xylene cyanole FF) were added. The reactions were then examined

on a 14% urea/acrylamide denaturing gel (acrylamide, 1 x TBE, 7 M Urea), which was run at 20 W for 50 minutes.

As no product was seen, the reactions were repeated with a 3-hour incubation at 37˚C. 1% SDS had to be added to the loading dye to denture the complex. The $MnCl_2$ concentration was raised to 10 mM and the $MgCl_2$ was removed from the buffer. The assay was also used on DS2, DS3, DS4, and DS8 RNAs. Different time courses, ranging to overnight, were performed. Finally, the best results were obtained using a 6-8 hour incubation at 37˚C with DS5 RNA, 5 mM $MnCl_2$, and no $MgCl_2$. These conditions were used to test different metals. Four metals (Mn^{2+}, Mg^{2+}, Ca^{2+}, and Zn^{2+}) were tested at 50, 500, and 5000 µM concentrations.

The time course experiment was also performed using the D15N mutant ScNob1.

1.11.4 Cleavage assays with potential co-factors

Dim1 is the protein responsible for di-methylation of the 20S RNA (Pulicherla et al, 2009). Initially, 0.5-3.0 µg of ScDim1 (purified by Anees Basha) and 160 µM *S*-Adenosyl methionine (SAM) were added to the above standard cleavage assay 2 hours prior to adding ScNob1. However, this yielded inconclusive results so the protocol was changed. 2 pmol of DS5 RNA were mixed with 3 µg of His- or GST-tagged ScDim1 and 160 µM SAM in buffer 1 (20 mM HEPES pH 7.5, 150 mM KCl, 4 mM Mg acetate, 5% v/v glycerol, and 1 mM β-mercaptoethanol) or buffer 2 (as buffer 1 but with 40 mM NH_4Cl instead of KCl as the salt) in a 10 µl volume. The buffers were designed based on previously published *in vitro* methylation assays (O'Farrell et al, 2006).

The reactions were then incubated at 37˚C for 2 hours. Then, 50 pmol of ScNob1 were added along with buffer 3 (as buffer 1 but with 5 mM $MnCl_2$ instead of Mg Acetate) in a further 10 µl, and incubated for a further 6 hours at 37˚C. The reaction was stopped by addition of equal volume (20 µl) denaturing dye (as above). As with the standard assay, the reactions were then analysed on a 14% denaturing urea/acrylamide gel (acrylamide, 1 x TBE, 7 M Urea).

As the results were still inconclusive, the same protocol was repeated but ScDim1 was removed from the reactions prior to addition of ScNob1. This was

achieved by the addition of 190 µl of splicing dilution buffer to the ScDim1 reaction after incubation and standard phenol/chloroform extraction and ethanol precipitation. The methylated RNA was resuspended in 5 µl of sterile water. The rest of the protocol was the same as previously.

1.11.5 Isothermal titration calorimetry (ITC)

ITC was used to measure the dissociation constants for the Nob1-D site interaction and to pinpoint the binding site. The protein for these experiments was purified as described previously except that the SEC buffer was changed to 20 mM HEPES (pH 7.5), 150 mM KCl, 5% v/v glycerol, 5 mM $MgCl_2$, and 1 mM TCEP (tris(2-carboxyethyl)phosphine). The RNAs were ordered from Sigma and were already HPLC-purified. Mutated sequences were designed not to have any secondary structure. This was tested for using the mfold server (http://mfold.rna.albany.edu). These were resuspended to a final 400 µM concentration in the SEC buffer.

To make sure that the protein and RNA buffers were matched, all binding partners were dialysed overnight at 4˚C into the same batch of 1-2 L SEC buffer. ScNob1 was dialysed using a Slide-A-Lyzer Dialysis Cassette (Thermo Scientific) according to manufacturer's instructions. The RNAs were dialysed in Micro Float-A-Lyzer dialysis units (Spectrum labs, molecular weight cut off of 0.1-0.5 kDa, 100-200 µl volume) also as suggested by the manufacturer.

The next day, the concentrations of the protein and RNAs were measured, using their specific extinction coefficients, on the NanoDrop (Thermo Scientific). The protein concentration was re-checked with a JASCO UV/Vis V-550 spectrometer to ensure accurate results. To obtain the necessary concentrations, the samples were diluted with the dialysis buffer. After dilution, the accurate concentrations were re-measured and recorded in the ITC200 software. All the experiments described below were performed using the MicroCal Auto-iTC200 system (GE Healthcare).

In the initial experiment, 10 µM ScNob1 was used in the cell with 150 µM DS6 (**Table 4.8** in the results section) RNA in the syringe. The samples were pipetted into a 96-well MicroCal Auto-iTC200 plate (GE Healthcare) and sealed. The plate tray was kept at 15˚C during the experiment. The titration was performed

over 16 injections (first one of 0.5 µl and 15 at 2.5 µl) at 25˚C with a reference power of 10. A 180 second gap was left between injections.

Then, the protein and RNA were reversed (10 µM DS6 RNA in the cell and 150 µM ScNob1 in the syringe); the reference power was reduced to 3 and temperatures of 37˚C and 5 ˚C as well as 25˚C were used. The first injection volume was reduced to 0.25 µl. The reversed titration at 5˚C with reference power at 3 was used for further experiments. The tray temperature was reduced to 10˚C. The stirring speed was 1000 rpm for all experiments. The RNA sequences used in the ITC can be seen in **Table 4.8** in the relevant results section.

Each set of experiments was accompanied by controls after the experimental set-up was chosen. The controls consisted of the same titrations but without one of the reactants. One set included buffer being injected into each of the RNAs separately and the second was ScNob1 being injected into buffer to eliminate dilution and buffer mismatch heat changes. When multiple experiments were performed in one run of the machine, the standard DS6 – ScNob1 experiment was performed before and after to ensure no changes, such as RNA degradation, occurred.

Once the experiments were finished, the data were processed in the ITC200 software. The experimental data were plotted together with the relevant controls. The baselines were adjusted to make sure all data were included. The peaks were integrated and the first point was removed (as the initial small injection is not correct). A linear regression was applied to the control data and subtracted from the experimental titration. Where possible, a one-site model was fitted to the binding curve. The molar ratio was set to 0.25 where required to reduce the dependency of the calculated values (i.e. where there was evidence of over-fitting).

All of the RNAs used were also separated on native PAGE to confirm the lack of secondary structures. This was performed on a 10% acrylamide gel in 0.5 x TBE. The gel was run at 180V at 4˚C for about 4 hours. The wells contained 2 µl of each 400 µM RNA in 50% v/v glycerol. Once finished, the gel was then soaked in 100 ml of 1x SYBR Safe (Invitrogen) stain solution for 5 minutes. It was then scanned using an XR Gel-doc system (Bio-Rad).

1.11.6 Surface Plasmon Resonance (SPR)

In order to estimate the dissociation constant for the Dim2-DS1 RNA interaction, SPR was performed. The His-tagged ScDim2ΔN51 was purified as described earlier. However, the size exclusion chromatography utilized a different buffer: 20 mM HEPES pH 7.5, 150 mM KCl, 50 µM EDTA, 0.05% v/v Tween20, and 1 mM TCEP. This buffer was used throughout the SPR experiments.

Initially, the experiments were performed at 25˚C. The protein was attached to an NTA chip (S Series Sensor chip, GE Healthcare) via nickel ions. The chip was placed in the BiaCore T200 machine (GE Healthcare). Channels 1 and 2 were washed with 15 µl of 50 mM NaOH and 90 µl of 350 mM EDTA. Channel 1 was then left as a reference. Meanwhile, channel 2 was loaded with nickel by a manual injection of 10 µl of 500 µM $NiSO_4$. Then, the 7.5-10 µl of 50 nM ScDim2ΔN51 were injected onto the surface of channel 2 (aiming for a response of 500-1000 units). Both channels were washed with buffer to stabilize the baseline.

The RNA was dialysed into the SPR buffer prior to the experiments. To check the surface activity, initially, 15 µl of 2 µM DS1 RNA were injected into both channels. Since the response was not at high as expected, a 75 µl injection was also attempted. Then, the RNA concentration was increased to 10 µM. A 60 µl injection at this concentration was used. The surface was then stripped with the same NaOH and EDTA wash as described earlier.

These initial experiments were then repeated at 10˚C to increase protein stability. Addition of 5 mM $MgCl_2$ made no difference to activity seen. A single cycle experiment was then performed at 4, 10, and 25˚C.

The surface was charged with nickel and protein as described earlier at each temperature using fresh protein. Five serial dilution of DS1 RNA were made (30, 15, 7.5, 3.75, and 1.875 µM). 15 µl of each concentration were injected (starting with the lowest concentration) onto the surface, separated by 30 seconds (i.e. 15 µl of buffer). The data were then analysed as suggested by the manufacturer.

1.11.7 ScDim2 Solubility screening

Attempts were made to create a soluble ScDim2ΔN51 – D site RNA complex. For this purpose, a grid screen was set up on two 24-well XtalQuest sitting

drop plates (MiTeGen). Each well contained 500 µl of the buffer being tested. Every condition had 5 mM $MgCl_2$, 5% v/v glycerol, and 1 mM DTT. On the horizontal axis, 12 buffers at different pHs were tested at 100 mM each (sodium citrate pH 5.5, sodium citrate pH 6.0, MES pH 6.0, Bis-Tris pH 6.5, HEPES pH 7.0, potassium phosphate pH 7.4, HEPES pH 7.5, Tris-HCl pH 7.5, Tris-HCl pH 8.0, Bis-Tris propane pH 8.0, Tris-HCl pH 8.5, and Bis-Tris propane pH 9.0). Vertically, different KCl concentrations were tested (50, 150, 300, and 600 mM).

On the pedestal in each of the 48 wells, 1 µl of 8.7 mg/ml (317 µM) ScDim2ΔN51 was mixed with 2 µl of well buffer and 1 µl of 381 µM (1.2 x protein concentration) D site RNA (DS1). The tops were sealed with clear tape and the plates were stored at 4˚C overnight. The next day, wells without precipitation were noted. These conditions were then tested on semi-analytical size exclusion chromatography as described earlier.

1.12 Dim2 Modelling

The potential structure of the eukaryotic Dim2 was modelled based on the solved archaeal structures. For this purpose, the *S. cerevisiae* protein sequence was entered into either the Phyre (http://www.sbg.bio.ic.ac.uk/phyre/ html/), Phyre 2 (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index), or Swiss-Model (http://swissmodel.expasy.org) server along with a multiple sequence alignment. The alignment consisted of some eukaryotic sequences (*S. cerevisiae*, *Schizosaccharomyces pombe*, *C. thermophilum*, *Caenorhabditis elegans*, *Wuchereria bancrofti*, *Drosophila melanogaster*, *Drosophila mojavensis*, *Acromyremex echinatior*, *Danio rerio*, *Mus musculus*, *Rattus norvegicus*, and *Homo sapiens*) and some archaeal ones (*Pyrococcus horikoshii*, *Aeropyrum pernix*, *Pyrolobus fumarii*, *Sulfolobus islandicus*, and *Methanococcus meripaludis*). The top hit for each one was used to create a model and these were then examined and compared to each other.

1.13 Calculating the Predicted Hydrodynamic Radii

The hydrodynamic radii (R_h) and radii of gyration (R_g) were calculated for existing solved structures using the HYDROPRO software (Ortega et al, 2011). For this purpose, the PDB files were edited to remove duplicate chains and additional atoms that were not a part of the protein (e.g. water molecules and ligands). The PDB files used were: BSA (ID: 4F5S), carbonic anhydrase (ID: 1V9E), and *P. horikoshii* Dim2 (ID: 2E3U).

An atomic level model using a shell calculation for each file was employed as advised by the creators of the program. The values were kept as the defaults apart from experimental variables (e.g. temperature or buffer viscosity), which were changed to reflect the conditions used in SEC experiments.

1.14 Electron Microscopy (EM)

ScNob1 was visualised using EM with and without D site RNA. For this purpose, a 5 nm carbon grid was charged in a SC7620 Mini Sputter Coater (Quorum) for 15 seconds. 4 µl of 900.0, 90.0, or 9.0 µg/ml ScNob1 were spotted onto the grid and left to sit for 2 minutes. Excess liquid was removed by blotting on filter paper and the grid was washed 3 times in sterile water. Then, the grid was washed 3 times in 2% w/v uranyl acetate. Then, about 5 µl of fresh 2% w/v uranyl acetate were spotted on and left for 5 minutes. The stain was removed and the grid was allowed to air dry for about 30 seconds. It was then examined using FEI F20 electron microscope (Tecnai). The protein with and without glycerol in the buffer was visualised.

For protein-RNA complexes, ScNob1 was first mixed with either 1.2, 4, or 10 times excess of DS1 RNA. The grids were prepared as described above but with 390 or 19.5 µg/ml (1 in 20 dilution) ScNob1 containing the excess of RNA.

1.15 Crystallisation

1.15.1 Crystallisation screening

Crystallisation screens were set up in 96-well MRC plates (Molecular Dimensions). The set up was performed by the Gryphon robot (Art Robbins) as suggested by the manufacturer. The drops were either made up of 0.1 µl of protein and 0.1 µl of well buffer or 0.2 µl of each if more protein was available. The well contained 60 µl of solution. Usually, a concentration of about 10 mg/ml of protein was used. After the pipetting, the plates were sealed using a XCS-384 pressure sealer (FluidX). The majority of plates were then stored at 16˚C (although some were also stored at 4˚C) and monitored for crystal growth. For the ScTsr1 constructs, all the initial screening was performed by Dr. Uma Jayachandran.

The commercial screens used included: PEG/Ion (Qiagen), Index (Hamilton), Magic I and II (Xtal Facility, MPI Martinsried), ProPlex (Molecular Dimensions) (Radaev et al, 2006), Midas (Molecular Dimensions) (Grimm et al, 2010), Structure 1 and 2 (Molecular Dimensions) (Jancarik & Kim, 1991), JCSG+ (Molecular Dimensions) (Newman et al, 2005), and Morpheus (Molecular Dimensions). Different combinations of screens were used for different proteins.

1.15.2 ScTsr1 crystallisation optimisation

1.15.2.1 ScTsr1Δ410-476

The initial hits occurred in wells A2, A7, and B3 in the Magic I screen using 8.7 mg/ml protein. Condition A2 consisted of 50 mM MES pH6.0, 10% w/v PEG 8000, and 0.2 M sodium malonate. A7 had 50 mM MES pH 6.0, 10% w/v PEG 3350, and 0.2 M sodium chloride while B3 was made up of 50 mM MES pH 6.0, 18% w/v PEG 8000, and 0.2 M calcium acetate. When these were set up in larger, XtalQuest sitting drop plates (MiTeGen), the first two conditions gave the largest crystals and were chosen for further optimisation. In the 24-well plates, the drops consisted of 1 µl of protein and 1 µl of well solution. The well contained 500 µl of the crystallisation mixture.

More 24-well plates were set up testing different variables. Initially, the PEG and salt concentrations were optimised using a grid screen. The conditions based on well A2 showed the best crystals. Different buffers were tested and Tris-HCl at pH 8.0 was then chosen. Some trials were also performed at pH 9.0.

Seeding was performed. For this purpose, the crystals grown at pH 8.0, 8% w/v PEG 8000, and 0.3 M sodium malonate were crushed with a needle. The fragments were mixed with 20 µl of well solution and placed in an eppendorf tube. The sample was centrifuged at 13000 rpm for 5 minutes. This stock and a 1 in 50 or 1 in 100 dilution were used to streak (using a cat whisker) drops containing 8.7 mg/ml protein and the well solution (50 mM Tris-HCl pH 8.0, 4% w/v PEG 8000, and 0.22-0.32 M sodium malonate).

To prevent the crystals sticking to the plate, a one-minute incubation of the plate with 5 µl of Gel Slick (Lonza) on the pedestal was tested.

Finally, the effects of additives were also checked. The pH 8.0, 8% w/v PEG 8000, and 0.3 M sodium malonate condition was used and supplemented with one of the following: 5 mM $MgSO_4$, 5 mM magnesium acetate, 5 or 10 mM $MnCl_2$, 5 or 10 mM $CaCl_2$, 5 mM calcium acetate, 50 µM zinc acetate, 5 or 10 mM glucose, 5 or 10 mM sucrose, 5 or 10% v/v glycerol, 5 or 10 mM GTP, 5 or 10 mM GDP, and 5 or 10 mM ATP. Afterwards, additional compounds were tested at pH 8.0 and 9.0. These included: 6-16 mM NAD, 6-16 mM TCEP, 6-16 mM DTT, 60-160 mM potassium sodium tartrate, 20-45 mM glycyl-glycyl-glycine, and 6-16 mM urea.

1.15.2.2 ScTsr1ΔN46Δ410-476

Again, Dr. Uma Jayachandran performed initial screening and some optimisation. I carried out optimisation around the best conditions identified, in 24-well plates using protein at 10 mg/ml in 8-14% w/v PEG 3350 and 0.18-0.30 M sodium malonate pH 6.0. Addition of buffers was tested including the following: 50 mM of MES pH 6.0, MES pH 6.5, HEPES pH 7.0, HEPES pH 7.5, Tris-HCl pH 8.0, and Tris-HCl pH 8.5. Conditions without buffer or with HEPES at pH 7.5 were further optimised. The protein concentration was lowered to 8 mg/ml.

The seleno-methionine labelled protein was also crystallised in the optimised conditions. As these crystals did not diffract as well, seeding using native crystals

was undertaken. For this purpose, the native crystals were treated in the same manner as described for the previous construct and the seeding was performed in the same manner. The protein was also re-screened at 10 and 5 mg/ml using the Hampton Index and Magic I screens as described previously.

1.15.3 Data collection

For data collection, the crystals were first harvested and placed in the well solution supplemented with 30% v/v glycerol (which acted as the cryo-protectant) or in stabilizing solution with glycerol (see later). They were then immediately harvested in nylon loops and flash-cooled in liquid nitrogen in ALS-style pucks (MiTeGen). They were stored in the liquid nitrogen until they were shipped to Diamond Light Source in Oxford where the data were collected remotely using the Generic Data Acquisition (GDA) client software. Initially, the crystals were quickly scanned by collecting 4 images 45˚ apart starting at 0 at 30% transmission with a 0.15 second exposure and 0.15˚ oscillation.

The data for the best ScTsr1ΔN46Δ410-476 crystal were then collected on the I03 beamline using a strategy calculated by EDNA analysis (Incardona et al, 2009) of the initial 4 images. Overall, 270˚ were collected starting at 78˚ with 0.15˚ rotation (i.e. over 1800 images). A 0.1 second exposure was used.

1.15.4 Data processing (native)

The 1800 images for the best ScTsr1ΔN46Δ410-476 crystal were processed using XDS software (Kabsch, 2010). All the jobs were run initially using a resolution range of 50.0 to 3.4 Å with a space group of 0. Friedel's law was set to 'true' and the rest of the settings were taken from the fast data processing files received from the Diamond Light Source (e.g. the origin of the beam, etc.).

The output file from the CORRECT program was examined. Then, the CCP4 suite of programs was used (Winn et al, 2011) for further processing. The XDS_ASCII.HKL file, which contained all of the corrected spot intensities, was entered into Combat along with the space group (16) and unit cell dimensions provided by XDS. This converted the data into .mtz file, which can be used to run SCALA. The intensities were scaled and merged using SCALA with the default

settings. Finally, Cad was used to convert the space group in that file from P222 (16) to $P2_12_12_1$ (19).

1.15.5 ScTsr1ΔN46Δ410-476 structure solution (phasing)

1.15.5.1 Molecular replacement

Solved structures of eIF5B (PDB ID: 1G7R) and SelB (PDB ID: 4ACA) were used as molecular replacement search models to obtain phases for the ScTsr1ΔN46Δ410-476 diffraction data. Initially, this was performed with the structures superimposed and the PDB files entered into Phaser (CCP4 suite) (McCoy et al, 2007; Winn et al, 2011) as separate ensembles along with the ScTsr1ΔN46Δ410-476 protein sequence and the .mtz file obtained from earlier processing. The program was told to search for two copies per asymmetric unit (default settings were used for the rest of the parameters). Initial data processing suggested that more than one protein molecule is present per asymmetric unit (see chapter 6). Thus, searching for 2 copies ensures that a more accurate model is found. This is balanced with processing time. Searching for more than two copies can take significantly longer as more structural elements need to be fitted.

The model sequences were then reduced and the side chains were removed using Chainsaw (CCP4 suite) (Stein, 2008; Winn et al, 2011) leaving poly-alanine sequences within the structures. These simplified models were again entered into Phaser (McCoy et al, 2007) with the ScTsr1ΔN46Δ410-476 data.

To further simplify the models, several truncations of each model protein were made to remove any non-conserved regions. For eIF5B these were at residue 443, 328, or 225 where anything following these was deleted from the PDB file. For SelB, the truncations were at 385, 269, and 178. Again, all six were entered as separate ensembles into Phaser (McCoy et al, 2007) with the data to obtain a solution.

Then, some restrictions were lifted to allow faster processing. The second truncation for each (328 for eIF5B and 269 for SelB) was rerun looking only for one copy per asymmetric unit and increasing the number of clashes allowed from 5 to 10.

Truncation 3 (225 for eIF5B and 178 for SelB) for both proteins was then used in a single ensemble (to include only the most conserved parts of the proteins). Again, only one copy per asymmetric unit was searched for.

The potential solution files were then examined in Coot (Emsley et al, 2010) to see if any structural features could be identified in the electron density maps.

1.15.5.2 Heavy atom soaking

For the heavy atom soaks, a stabilising solution for the crystals was identified by harvesting crystals and placing them in higher PEG and sodium malonate conditions overnight and making sure no dissolving occurred. 14% w/v PEG 3350 and 0.3 M sodium malonate were chosen. The heavy compounds used were thimerosal, gold (I) potassium cyanide, lead (II) acetate trihydrate, samarium (III) acetate hydrate, and potassium tetranitroplatinate (II). These were made up in sterile water to a final concentration of 30 mM. The 24-well plate was prepared by placing 4.5 µl of stabilising solution on each pedestal and 500 µl of the same solution in each well. 0.5 µl of each heavy atom stock at 10, 1, or 0.1 mM concentration was added to each pedestal to give a final volume of 5 µl.

Crystals were then harvested from 10-12% w/v PEG 3350 and 0.22-0.30 M sodium malonate and placed in the 5 µl drop of the stabilising solution supplemented with the heavy atoms. About 3 crystals were placed on each pedestal and sealed. They were left to incubate at 16˚C for 3 hours. Then, the crystals were harvested and moved to a cryo-protectant solution (back-soaked), which consisted of the stabilising solution supplemented with 30% v/v glycerol. They were then immediately picked up again and flash-cooled in liquid nitrogen and stored until data collection at the Diamond Light Source in Oxford.

As for the native crystals, data were collected remotely. Initially, to detect if any heavy atoms were present and to determine what wavelength to use for data collection, a fluorescence scan was performed. Heavy atoms absorb X-rays at distinct energies and release them as fluorescence. The required heavy atom was selected along with edge L3 and an energy scan was performed, as advised by the Diamond Light Source user guidelines. If the heavy atom was present in the crystal, the peak and inflection energies at which the fluorescence occurred were recorded.

Then, initial crystal scans were collected as normal (4 images as described earlier) except that the wavelength was changed to either the peak or inflection energy. EDNA (Incardona et al, 2009) was used to calculate a strategy to collect full data sets. **Table 2.10** lists the different strategies used for anomalous dataset collection. Afterwards, the data were processed as described above except that Friedel's law was set to 'False' in XDS.

Table 2.10: The strategies used for anomalous data collection.
The crystal name corresponds to the heavy atom present and whether data were collected at the peak or inflection energy. Some data were also collected at high remote energy (hrm), which is reached after the peak fluorescence energy. The Diamond Light Source beamlines used for each set of data are also shown.

Crystal	Beam -line	Energy (eV)	Omega start	Rotation	Number of images	Exposure (seconds)	Resolution (Å)	Trans-mission (%)
Hg 1 - peak	I03	12298.5	207˚	0.2˚	900	0.2	4.0	30
Hg 2 - peak	I03	12298.5	0˚	0.15˚	2400	0.15	4.0	30
Hg 2 - inflection	I03	12282.5	0˚	0.15˚	2400	0.15	4.0	20
Hg 3 – peak	I24	12311.7	289˚	0.2˚	450	0.2	4.0	12
	I24	12311.7	352˚	0.2˚	450	0.2	4.0	20
	I24	12311.7	352˚	0.2˚	450	0.2	4.0	50
Hg 3 - inflection	I24	12293.0	352˚	0.2˚	450	0.2	4.0	30
	I24	12293.0	262˚	0.2˚	450	0.2	4.0	30
Hg 4 - hrm	I24	12315.0	289˚	0.2	550	0.2	4.0	30
Hg 4 - peak	I24	12311.7	289˚	0.2	550	0.2	4.0	30
	I24	12311.7	199˚	0.2	550	0.2	4.0	30
Pb 1 - peak	I03	13050.0	0˚	0.15˚	2400	0.15	4.0	20
Sm 1 - peak	I03	6716.0	0˚	0.15˚	2400	0.15	4.0	20
Sm 1 - peak	I03	6716.0	0˚	0.15˚	2400	0.15	4.0	20

After XDS, SHELXC/D/E (Sheldrick, 2010) was used as recommended by the documentation to see if any anomalous signal was present to allow calculation of a substructure for phasing.

1.15.5.3 Seleno-Methionine protein

The seleno-methionine ScTsr1ΔN46Δ410-476 was purified and crystallised as described earlier. As for the heavy atom-soaked crystals, initially, a fluorescence screen was performed to identify the peak (12662 eV) and inflection (12660 eV) energies. Two datasets were collected for the best diffracting crystal (about 10 Å resolution) as listed in **Table 2.11**.

Table 2.11: Datasets collected using the best seleno-methionine protein crystal.

Dataset	Beam -line	Energy (eV)	Omega start	Rotation	Number of images	Exposure (seconds)	Resolution (Å)	Trans-mission (%)
Peak	I04	12662	0˚	0.1˚	1800	0.1	4.0	30
Inflection	I04	12660	0˚	0.1˚	1800	0.1	4.0	30

The datasets were processed in XDS as described for the heavy atom-derived crystals. Again, SHELX (Sheldrick, 2010) was used as recommended by the documentation to see if sufficient anomalous signal was present to allow for phasing.

1.16 Standard Laboratory Protocols

1.16.1 SDS-PAGE

Samples were prepared by addition of equal volume of 2 x sample buffer (100 mM Tris-HCl at pH 6.8, 10% v/v β-mercaptoethanol, 4% w/v SDS, 0.2% w/v bromophenol blue, 20% v/v glycerol). They were then boiled at 95˚C for 5 minutes, spun down, and loaded onto a denaturing SDS/acrylamide gel of the appropriate percentage. The percentage of acrylamide was chosen based on the size of the protein to be visualised. For this project, 10 or 15% gels were utilised. The gels were run at 180V for 1 hour. They were then washed with water and stained with diluted Coomassie (10% v/v ethanol, 5% v/v acetic acid, 0.001% w/v Coomassie) overnight or silver stained as described later.

1.16.2 MALDI-ToF

1.16.2.1 Sample Preparation

The samples to be analysed by MALDI-ToF were separated using standard SDS-PAGE and stained with diluted Coomassie overnight. The gel was then washed for a day with clean water. The bands of interest were excised and incubated in 300 µl of 200 mM ammonium bicarbonate in 50% v/v acetonitrile (buffer 1) at room temperature for 30 minutes. This step was performed twice to remove the SDS. The gel fragments were then incubated in 300 µl of buffer 2 (20 mM DTT, and 200 mM ammonium bicarbonate in 50% v/v acetonitrile) at 32˚C for 1 hour to reduce the protein.

The buffer was removed and the gel was washed 3 times with 300 µl of buffer 1. This was followed by a 20-minute incubation in the dark (room temperature) in buffer 3 (50 mM iodoacetamide, and 200 mM ammonium bicarbonate in 50% v/v acetonitrile) to alkylate the cysteines. The gel pieces were then washed 3 times with buffer 4 (20 mM ammonium bicarbonate in 50% v/v acetonitrile) and spun down at 13,000 rpm for 2 minutes.

Finally, the samples were covered with 200 µl of acetonitrile (until they turned white). The solution was decanted and the gel pieces were allowed to air dry. 30 µl of trypsin (Sigma) at 13.3 µg/ml in 50 mM ammonium bicarbonate were added and the gel pieces were allowed to swell in the solution at 4˚C. The tubes were sealed with parafilm and incubated at 32˚C for 16-24 hours. They were then stored at -20˚C until needed.

1.16.2.2 Matrix Preparation

For peptide analysis, 10 mg of a-cyano-4-hydroxyconnamic acid (CHCA) were weighed out and mixed with 400 µl of sterile water. 100 µl of 3% v/v trifluoroacetic acid (TFA) were added and the sample was mixed. Then, 500 µl of acetonitrile were added followed by more mixing. This was spun down at 5,000 rpm for 1 minute. The supernatant was used for spotting onto the plate.

1.16.2.3 Sample analysis

0.5 µl of peptide sample and 0.5 µl of CHCA matrix were spotted unto a gold or a stainless steel 100-sample plate (Applied Biosystems). The plate was allowed to air dry and then it was inserted into the Voyager DE-STR (Applied Biosystems) machine. Data were collected as advised by the manufacturer.

Once the peaks were obtained, they were processed in Data Explorer (Applied Biosystems), again, according to manufacturer's recommendations. The size of the fragments was calibrated based on visible trypsin peaks. Finally, the Mascot search engine (http://www.matrixscience.com) was used to identify the protein fragments based on known protein sequences.

1.16.3 Silver staining

Pre-cast 4-20% gradient gels (Bio-Rad) or standard gels were loaded and run at 180V for 35 minutes. They were then fixed in 50 ml 50% v/v methanol three times (10 minutes each). This was followed by at least 10 minutes of rehydration in 5% v/v methanol. The gels were then rinsed 3 times with water and soaked in DTT at 5 µg/ml for 20 minutes. The DTT was removed and the gels were soaked in 0.1% w/v silver nitrate for 20 minutes. This was followed by a quick rinse with 100 ml water and a change into a clean container.

The gels were rinsed twice with 50 ml developer solution (7.5 g Na_2CO_3 and 0.02% v/v formaldehyde in 250 ml water). Then, 50 ml of fresh developer was added to each gel and the exposure was carefully monitored. Once satisfactorily exposed, 5 ml of 2.3 M citric acid were added to stop the reaction. The gels were left to mix for 10 further minutes and then rinsed with water.

1.16.4 RNA native gels

Native gels were used for EMSAs and secondary structure analysis of RNAs. The 20 x 20 cm gels consisted of the appropriate percentage of acrylamide and 0.5 x TBE buffer. The samples to be loaded onto the gel were mixed with 6 x non-denaturing dye (0.25% w/v bromophenol blue, 0.25% w/v xylene cyanole FF, and 50% v/v glycerol) and kept on ice or at room temperature. After loading, the gel was run in 0.5 x TBE buffer at 200V at 4˚C or at room temperature until the dye was halfway down the gel (usually between 2 and 4 hours depending on the percentage of acrylamide).

1.16.5 RNA gel purification

RNAs were separated on a 12-16% denaturing urea/acrylamide gel (acrylamide, 1 x TBE, 7 M Urea). The 20 x 20 cm gel was run in 1 x TBE at 20 W for 45 minutes to separate the RNAs. The gel was then wrapped in cling film and placed on a Fluor-coated TLC plate (Ambion). The RNAs were visualised using UV shadowing under short wave (254 nm) light and marked on the cling film. The bands were excised and incubated in 400 µl of elution buffer (300 mM NaCl, 60 mM NaOAc at pH 5.2, and 0.2% w/v SDS) overnight. The following day, they were

phenol/chloroform extracted and ethanol precipitated as described below. The RNA was resuspended in the required amount of sterile water.

1.16.6 Phenol/Chloroform extraction

An equal volume of phenol/chloroform/IAA (25:24:1, Ambion) was added to the sample. The tube was vortexed twice. The samples were spun down at 13,000 rpm for 1 minute at room temperature. The aqueous phase (top layer) was carefully collected (without touching the bottom phase) and placed in a clean tube.

1.16.7 Ethanol Precipitation

Ethanol (100%) was added to the tubes (usually 1 ml to give a final concentration of about 70% v/v ethanol). These were vortexed and placed at -20˚C for 1 hour. They were then centrifuged at 13,000 rpm at 4 ˚C for 15 minutes. The liquid was discarded and the pellet was washed with 300 µl of 80% v/v ethanol. The tubes were spun again at 13,000 rpm at 4˚C for 1 minute. The liquid was discarded and the pellets were allowed to dry. They were then resuspended in the appropriate amount of buffer or water.

1.16.8 Limited Proteolysis

For limited proteolysis, 10 µl of the protein of interest (at a concentration of 0.6 mg/ml) were mixed with 3 µl of a dilution of a protease. The proteases used were trypsin, elastase, chymotrypsin, GluC, and subtilisin (Sigma). There were 3 dilutions used for each (0.1, 0.01, and 0.001 mg/ml) in 20 mM HEPES pH 7.5, 50 mM NaCl, and 10 mM $MgSO_4$. The reactions were incubated on ice for 30 minutes and then 2 µl of SDS-loading dye were added to stop the cleavage. The samples were analysed on 12.5% gels using standard SDS-PAGE.

1.16.9 PCR protocols

The following PCR protocols were used throughout the project:

Gene/plasmid amplification:

Initial denaturation:	95°C for 3 minutes
38 cycles:	95°C for 30 seconds

	65°C for 30 seconds
	72°C for 15–30 seconds/kb
Final extension:	72°C for 5 minutes

RT-PCR protocol:

First strand synthesis:	58˚C for 15 minutes
Initial denaturation:	95°C for 4 minutes
40 cycles:	95°C for 30 seconds
	60°C for 30 seconds
	68°C for 10 seconds
Final extension:	68°C for 5 minutes

1.16.10 Transformation protocols

1.16.10.1 Chemically-competent cells

The appropriate volume of DNA (usually between 1-3 µl of plasmid extracted using the MiniPrep kit) was added to 50 µl of aliquoted cells. The mixture was stored on ice for 30 minutes. Then, heat shock was applied by moving the cells to 37˚C for 30 seconds and immediately following placing them back on ice for a further 5 minutes. 200 µl of SOC medium were added and the cells were shaken at 37˚C for 1 hour. They were then plated on agar supplemented with the appropriate antibiotics.

1.16.10.2 Electro-competent cells

Again, DNA was added (up to 2µl) to the aliquoted cells. The mixture was placed on ice in a 0.2 cm gap Gene Pulser cuvette (Bio-Rad). Electroporation was performed in MicroPulser Electroporator (Bio-Rad) according to the manufacturer's instructions. 200 µl of SOC medium were added and the cells were shaken at 37˚C for 1 hour in the same manner as the chemically-competent cells. Again, these were plated on agar containing the appropriate antibiotics.

3 CHARACTERISATION OF NOB1

3.1 Introduction

As previously mentioned, Nob1 is the nuclease responsible for D site cleavage and is conserved in yeast and humans. This protein consists of two domains: a PIN domain and a zinc ribbon. The PIN domain is normally associated with magnesium-dependent exonucleases (Arcus et al, 2004). A single mutation in this domain (D15N in the yeast or D10N in the human protein) was shown to abolish the nuclease activity of Nob1 (Lebaron et al, 2012). The zinc ribbon domain is thought to be necessary for RNA interactions (Granneman et al, 2010).

During ribosome assembly, Nob1 localises to the nucleus and the cytoplasm (Fatica et al, 2003). Its activity appears to be inhibited until the final stages of pre-18S maturation are reached (Granneman et al, 2010). It is possible that it forms a functional tetramer *in vitro* (Lamanna & Karbstein, 2009) although no conclusive evidence is available to date that this is the case *in vivo*.

Nob1 has been shown to interact with multiple proteins, including Prp43, Dim2, and Fap7 (Hellmich et al, 2013; Vanrobays et al, 2008; Woolls et al, 2011). In archaea, the interaction with Fap7 blocks D site cleavage *in vitro* (Hellmich et al, 2013). However, the cleavage cannot occur in yeast *in vivo* without Fap7 expression (Granneman et al, 2005).

An NMR structure of the individual domains of the *P. horikoshii* Nob1 protein is available. The domains are linked by a flexible region that could not be mapped in a single conformation (Veith et al, 2012). Also, the archaeal protein is much shorter than its eukaryotic counterpart. Due to the complexity of the ribosome biogenesis in eukaryotes, Nob1 has to be separated from the other components in the pathway in order to understand its specific function and the mechanism by which it is performed.

During the course of this project, eukaryotic recombinant Nob1 from *C. thermophilum* (Ct), *S. cerevisiae* (Sc), and *H. sapiens* (Hs) was expressed, purified, and characterised. The results of the characterisation are presented in this chapter.

3.2 Cloning and Expression Trials

In order to study Nob1 function, different constructs for protein expression were created. These included yeast and human versions of the protein, as shown in **Table 3.1**. The ScNob1 full-length and the D15N mutant sequences in the pSL6 plasmid were obtained from the Tollervey lab. The expression trials for these constructs were repeated and can be seen in **Figure 3.1**. Two media were tested (Superbroth and 2xTY) and 2xTY was chosen for all following experiments. The expression for the D15N mutant seemed to be as good as for the wild type (WT) protein. However, its ability to interact with nickel beads was decreased.

Table 3.1: List of Nob1 constructs.
The top four constructs were present prior to the start of the project while the rest were created during.

Insert	Organism	Plasmid	Comments
Nob1 Full-length (FL)	*S. cerevisiae*	pSL6	Expressed, soluble – only in SE1 cells
Nob1 1-125 (PIN domain)	*S. cerevisiae*	pEC-KHC, pEC-KGC	Expressed, Majority insoluble
Nob1 264-459 (Zn ribbon)	*S. cerevisiae*	pEC-KHC, pEC-KGC	Expressed, Majority insoluble
Nob1 D15N	*S. cerevisiae*	pSL6	Expressed, soluble
Nob1 Full-length (FL)	*S. cerevisiae*	pET28α	Expressed, Soluble
Nob1 FL	*H. sapiens*	pEC-KHC	Expressed, Soluble; Degradation seen
Nob1 D10N	*H. sapiens*	pEC-KHC	Expressed, Soluble; Degradation seen
Nob1 FL	*H. sapiens*	pET28α	Expressed, Soluble; No degradation
Nob1 D10N	*H. sapiens*	pET28α	Expressed, Soluble; No degradation

Other constructs were successfully cloned and the sequences were confirmed to be correct by Sanger sequencing prior to further use. The human protein sequence was originally obtained in the pET100 plasmid from Dr. Nick Watkins (University of Newcastle) and then re-cloned into pEC-KHC (see **Appendix II**) and pET28α. **Figure 3.2** shows the results of the expression tests for these constructs. These results are also summarised in the table. The expression of the full-length yeast protein was not as good as using the original clone in the pSL6 vector so it was not used after optimisation of purification.

The expression of HsNob1 was good. However, putative degradation products could be seen with the His-tag on the N-terminus, suggesting that truncations were occurring on the C-terminus. These were not seen when the His-tag

was moved to the C-terminus as the degradation products now lacked the tag and did not bind to the nickel resin. The D10N mutant (like the D15N yeast mutant) showed a decreased ability to bind to nickel beads, although moving the tag to the C-terminus improved the interaction.

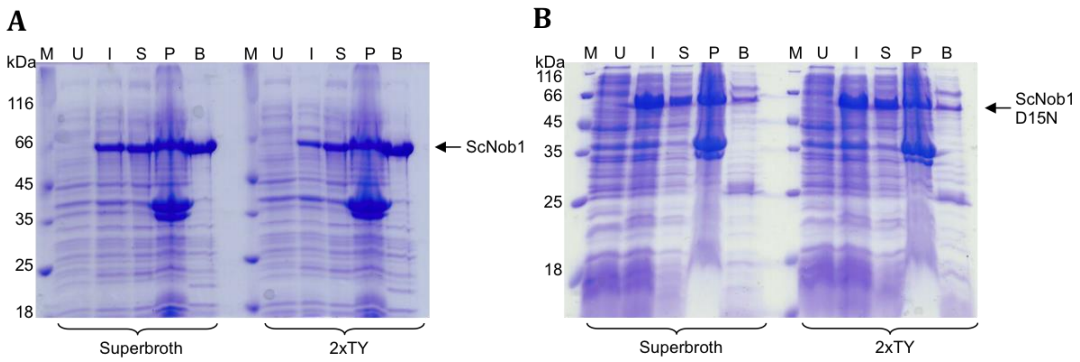


Figure 3.1: Expression trials for His-tagged wild type (WT) ScNob1 and the D15N mutant obtained from the Tollervey lab.

The two proteins were in the pSL6 plasmid (with a non-cleavable C-terminal His-tag) and were expressed in SE1 cells in either Superbroth or 2xTY medium. The different lanes correspond to the different fractions: uninduced (U), induced (I), soluble (S), insoluble/pellet (P), and the nickel bead-bound (B) fraction. The molecular weight markers (M) are also shown on each gel. The band corresponding to the correct size is indicated with an arrow. The same labels are used in all following expression trials.

A) Expression of the WT ScNob1 in the two media. Both show large amounts of soluble protein. However, a larger portion of the soluble protein in 2xTY bound to the beads so this medium was chosen for further preparations.

B) Expression of the D15N mutant version of the ScNob1 protein in the different media. Again, expression levels between the media are comparable. Like the WT, all following cultures were grown in 2xTY. The mutant protein shows a reduced ability to bind to the nickel beads compared to the WT even though the expression level is good.

During the project, Dr. Atlanta Cook also cloned ScNob1 "loop-out" mutants. These lacked different regions of the large flexible loop connecting the PIN domain to the zinc ribbon. These clones are listed in **Table 3.2**. As for the other constructs, expression tests were performed and the results can be seen in **Figure 3.3**. All the mutants were expressed in BL21 (DE3) cells and showed better enrichment on nickel beads than the WT protein.

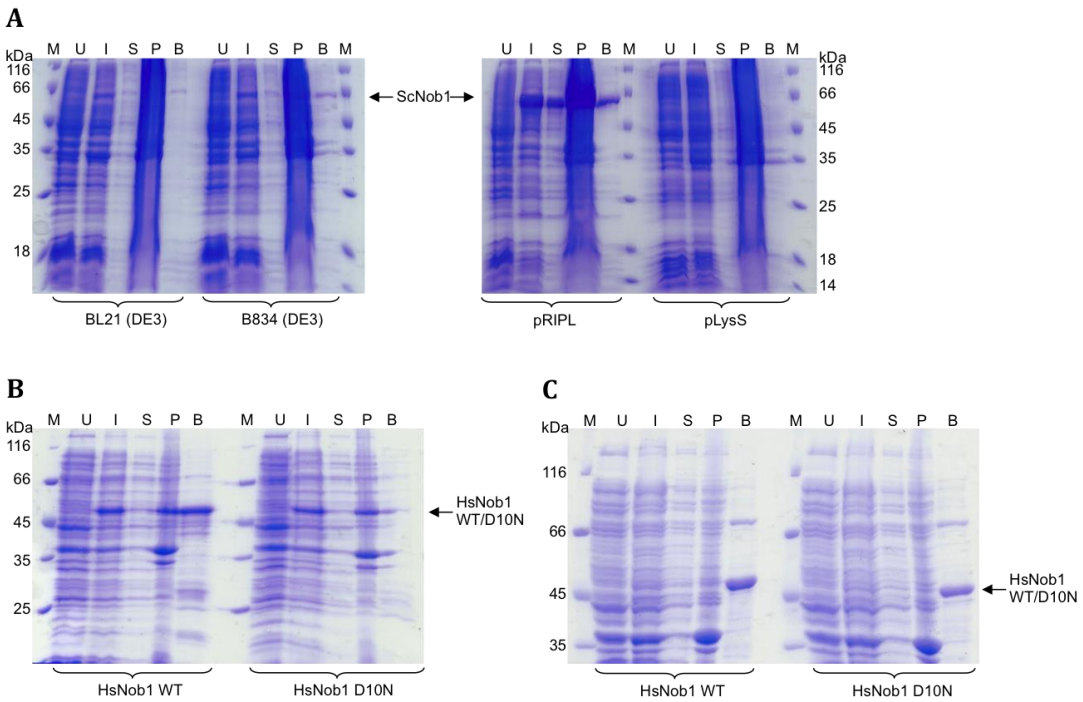


Figure 3.2: Expression trials for the different Nob1 constructs created during this project.
The lane labelling is the same as for the previous figure.

A) Expression of ScNob1 cloned into the pET28α vector in different cells lines. The cell line used is shown under each set of fractions. Faint bands at the correct size can be seen in bead-bound fractions using BL21 (DE3) and B834 (DE3) strains. The BL21 (DE3) pRIPL cell line shows a much higher level of expression. No protein can be observed in any of the lanes for the BL21 (DE3) pLysS sample. In following experiments, either the pRIPL or the BL21 (DE3) cell lines were used for expression. After the purification conditions were established, the pSL6 construct was used instead.

B) Expression of HsNob1 (and the D10N mutant) cloned into the pEC-KHC plasmid. These clones have a cleavable His-tag on the N-terminus. The tests were performed in the BL21 (DE3) cell line. The WT protein (left) shows some soluble expression, which is enriched in the bead-bound fraction. The D10N mutant (right) shows a similar level in the induced fraction as the WT but it does not bind to the nickel resin as efficiently, as shown by the lack of enrichment in the bead-bound fraction. The WT protein shows what might be degradation products in the bead-bound lane as there are multiple bands, below the full-length protein, which bound to the nickel beads.

C) Expression of HsNob1 (and the D10N mutant) cloned into the pET28α plasmid where the His-tag is on the C-terminus and is non-cleavable. The level of WT protein in the induced fraction is less than for the other construct. However, the level is still enriched on nickel beads and fewer degradation products are seen. The D10N mutant shows better enrichment on the beads. The N- and C-terminally tagged proteins were tested in purification as shown later.

Table 3.2: List of "loop-out" mutants cloned by Dr. Atlanta Cook.

Insert	Organism	Plasmid	Comments
Nob1ΔL1 (Δ151-159)	*S. cerevisiae*	pET28α	Expressed, Soluble
Nob1ΔL2 (Δ151-169)	*S. cerevisiae*	pET28α	Expressed, Soluble
Nob1ΔL3 (Δ151-176)	*S. cerevisiae*	pET28α	Expressed, Soluble
Nob1ΔL4 (Δ141-176)	*S. cerevisiae*	pET28α	Expressed, Soluble
Nob1ΔL5 (Δ141-159)	*S. cerevisiae*	pET28α	Expressed, Soluble

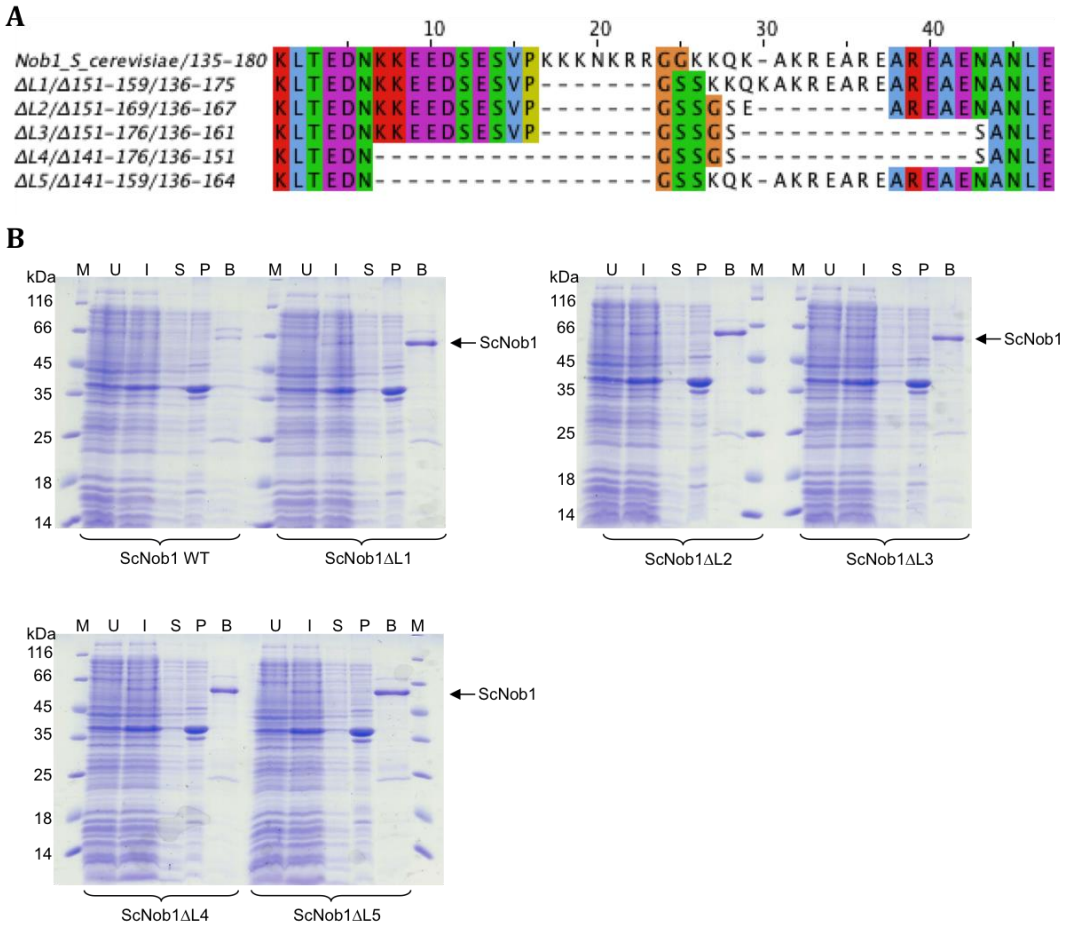


Figure 3.3: Alignment and expression of the ScNob1 loop-out mutants created by Dr. Atlanta Cook.

A) A local sequence alignment (showing residues 135-180) of the full-length ScNob1 and the loop-out mutants.

B) The lanes are labelled as before and the construct being tested is shown below each set of samples. All the mutants and the WT protein were in the pET28α plasmid containing a non-cleavable C-terminal His-tag. The expression was performed in BL21 (DE3) cells. As seen before with this construct, the WT protein shows a weak band in the bead-bound fraction. All of the mutants (ScNob1ΔL1-L5) show an increased level of protein in the bead-bound fraction compared to the WT suggesting that ScNob1 is stable and soluble without parts of the flexible loop.

Finally, Nob1 from *C. thermophilum* was ordered as a codon optimised construct in the pMK-RQ plasmid from GeneArt. The expression was good as shown in **Figure 3.4**. The BL21 (DE3) cell line was chosen for further cultures.

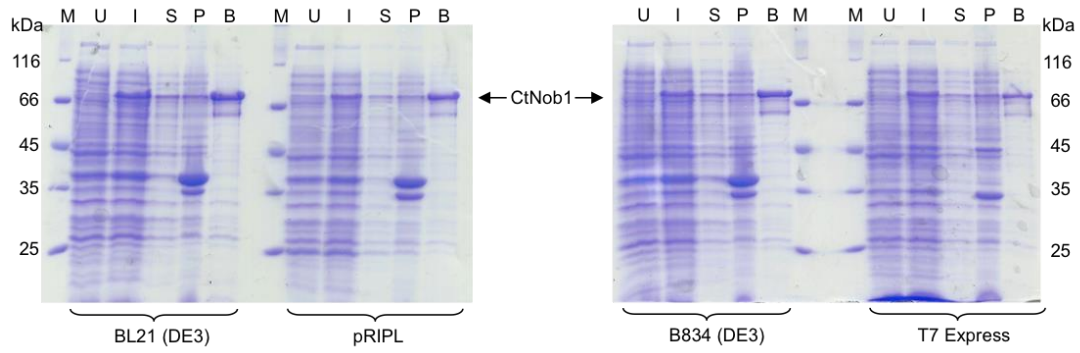


Figure 3.4: Expression trials for CtNob1 in different cell lines.
Again, the labels are the same as before and the cell lines are shown below each set of samples. The protein was expressed and soluble in every cell line tested. The T7 Express cells show the least enrichment on the beads. The BL21 (DE3) cell line appears to have the highest expression level and was, thus, chosen for all following preparations.

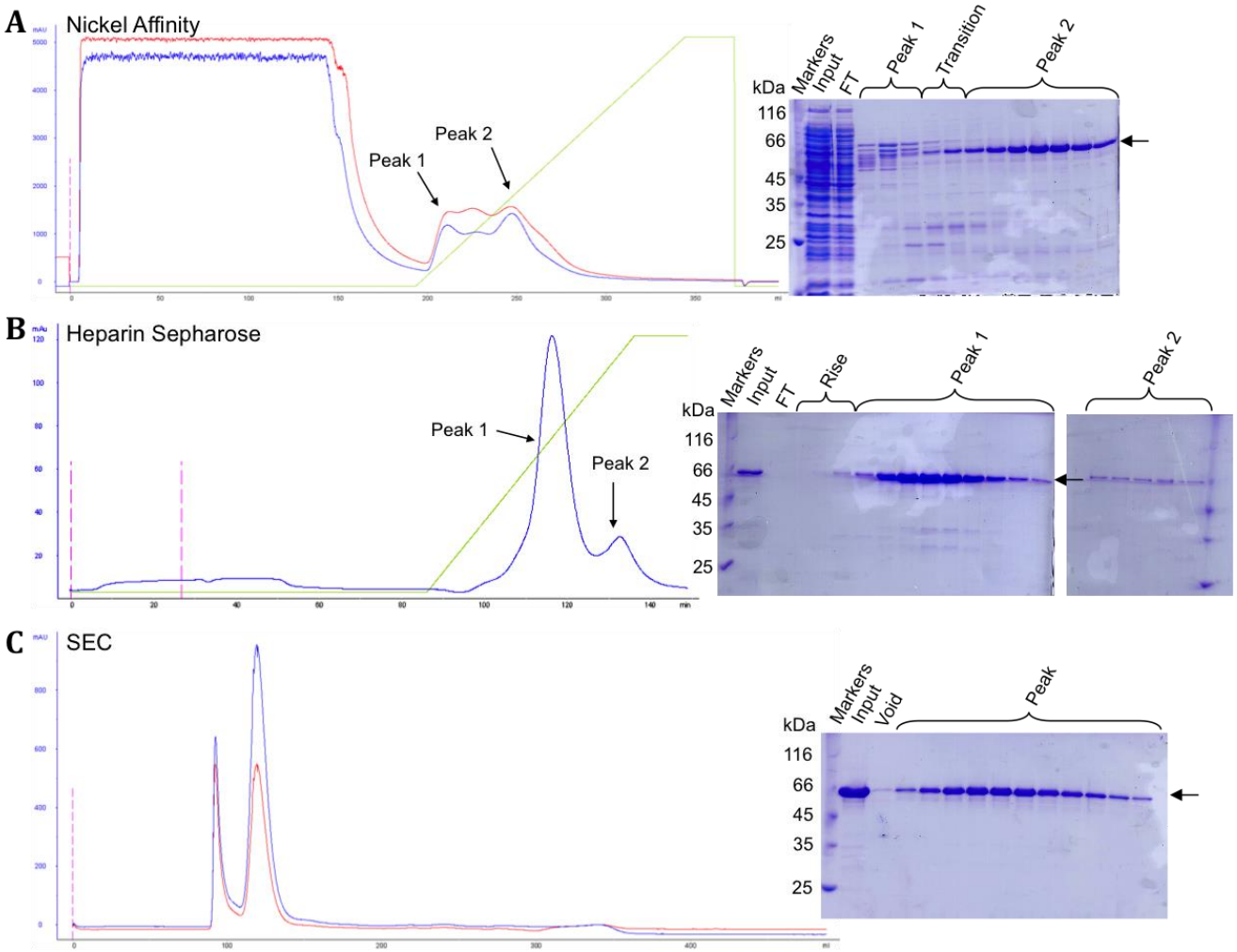
3.3 Protein Purification

3.3.1 ScNob1 – full-length

A three-step purification protocol for ScNob1 had been previously established involving affinity capture (nickel resin binding the His-tagged protein) followed by heparin sepharose and size exclusion chromatography (SEC). All steps included buffers supplemented with glycerol (10% for the first two steps and 5% for the third). Relatively pure protein was obtained (**Figure 3.5**). However, as seen in **Figure 3.5C**, contaminants, including RNases, were still present at the SEC step. The protocol was further optimized by removing glycerol from the buffers and using a cation exchange step instead of heparin sepharose, which improved the quality of the protein, especially at the SEC step (**Figure 3.6**). The protein was predicted to bind to the cation exchange column based on its calculated isoelectric point (pI of 8.5). After optimisation, about 15 mg of protein were obtained from 2 litres of bacterial culture.

Figure 3.5: Initial purification of ScNob1.

All chromatograms show absorbance at 280 nm and 260 nm (when available) in blue and red respectively. The pink dotted line corresponds to the injection of the sample. The appropriate elution gradient is shown as a green line. The protein of interest is indicated on the gels with an arrow.



A) Nickel resin purification. The green line shows the imidazole gradient (20-500 mM) used for elution. On the right, an SDS-PAGE gel of the different fractions is shown. ScNob1 runs just below 66 kDa and is the most visible band. Anything that came off the column prior to the gradient was considered to be flowthrough (FT).

B) Heparin sepharose column purification. The green line shows the salt gradient (200-1000 mM). Again, SDS-PAGE gels of the fractions are shown with the band corresponding to ScNob1 marked with an arrow.

C) SEC of the ScNob1 sample. The first peak corresponds to the void volume (i.e. FT) and shows that some contamination is present (possibly RNA contamination as it displays relatively high absorbance at 260 nm compared to 280 nm).

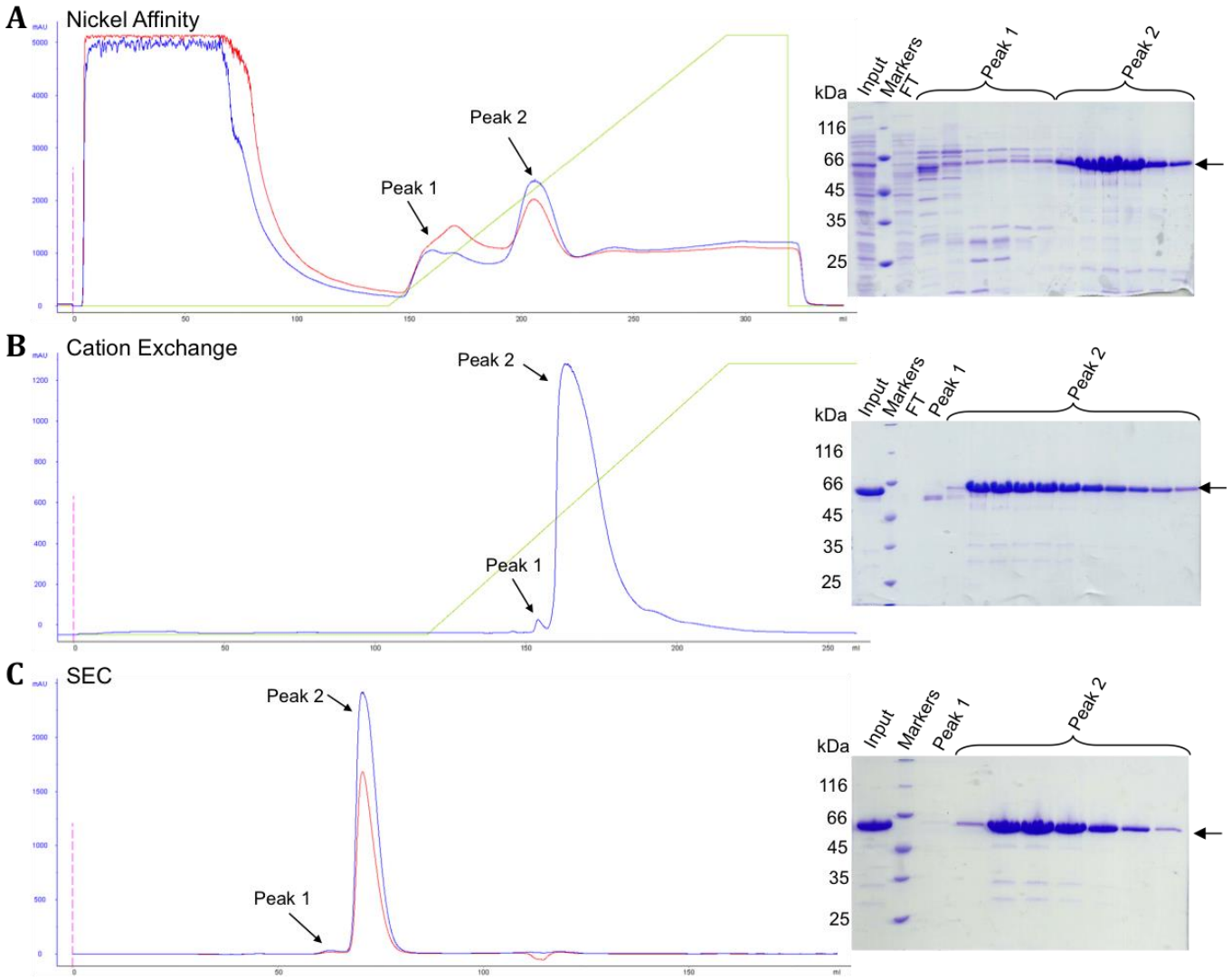


Figure 3.6: Optimised purification of ScNob1.
All the labelling and colour schemes are the same as on the previous figure. No glycerol was used in the buffers and the second step was changed.
A) As before, the first step consisted of nickel affinity purification. The profile appears different, as the buffers used this time did not contain glycerol.
B) The second step was changed to cation exchange i.e. Mono S column. The elution profile can be seen. The gradient ranged from 100 to 1000 mM salt. The majority of the contaminants precipitated out of solution during dialysis into 100 mM NaCl prior to this step. The gel on the right shows that the protein coming off during the gradient is relatively clean.
C) The final step was, again, SEC. No FT in the void volume is seen since the input is cleaner. The extra bands seen on the gel appear to be degradation products and not contamination.

3.3.2 CtNob1 – full-length

CtNob1 was first purified using the same starting protocol as ScNob1. However, as with ScNob1, contaminants were still present after size exclusion chromatography (**Figure 3.7**). Also as with ScNob1, a similar improvement in purity was observed when an anion exchange step was included after initial affinity capture (**Figure 3.8**). An anion exchange step, instead of cation exchange, was used due to the predicted pI (6.1). About 11 mg of protein could be obtained from 4 litres of culture.

3.3.3 HsNob1 – full-length

Initially, the N-terminally tagged protein (expressed from the pEC-KHC plasmid) was subjected to purification. As with the protein from the other two organisms, nickel affinity was used as the first step. Then, the tag was cleaved and the protein was loaded onto an anion exchange column (pI of 6.6). However, as shown in **Figure 3.9**, there were many degradation products present that could not be separated using these two steps. Therefore, no further attempts were made using this construct.

The protein was re-cloned to contain a non-cleavable C-terminal His-tag to prevent the binding of degradation products to nickel resin. The purification was performed as above but an ATP wash was included in the first step to remove any associated chaperones. After the anion exchange, a SEC step was performed (**Figure 3.10**). The HsNob1 obtained was not completely pure and only about 30 µg of protein were obtained from 4 litres of bacterial culture.

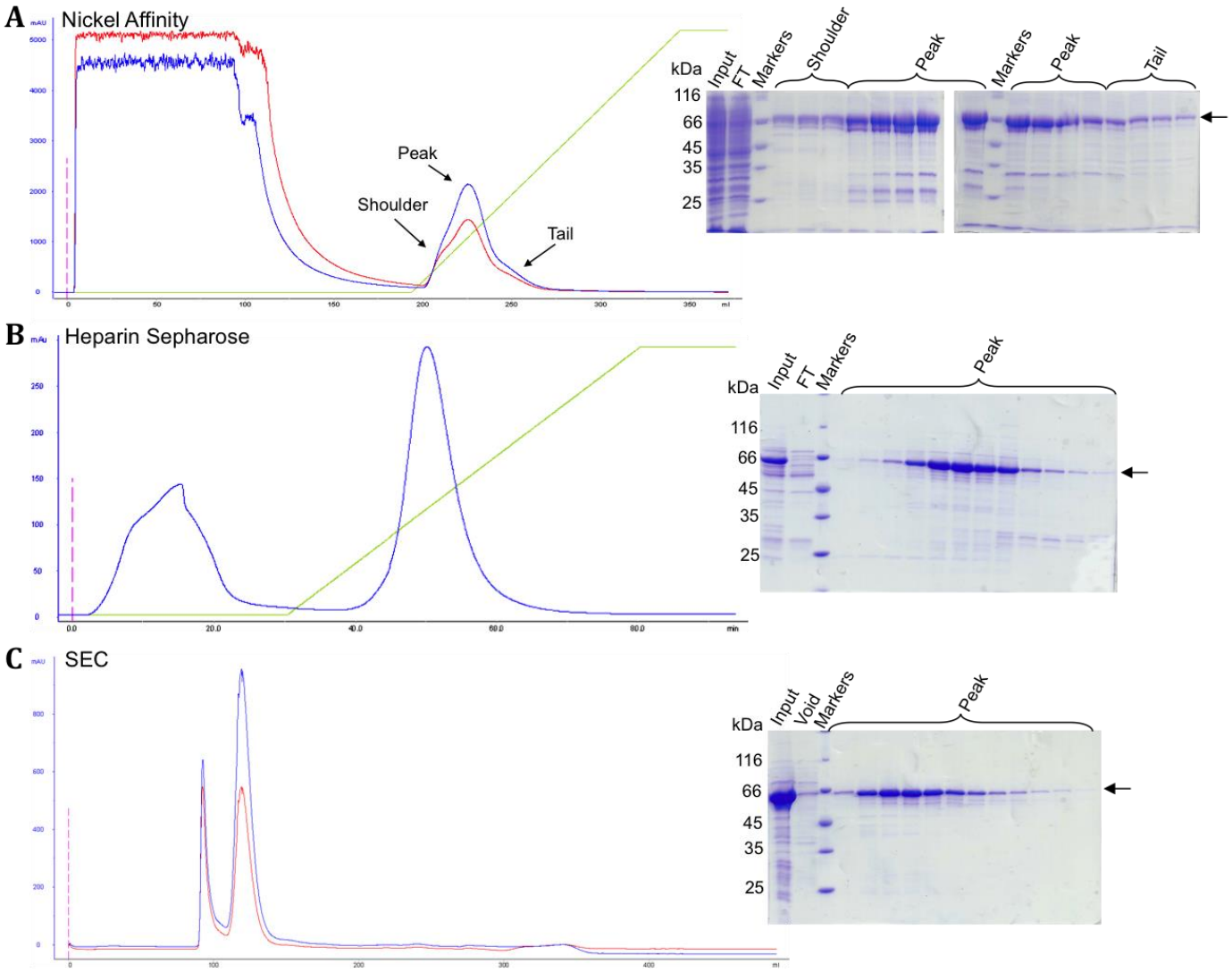


Figure 3.7: Initial purification of CtNob1.

All the labelling and colour schemes are the same as on previous figures. The protocol is the same as, initially, for the *S. cerevisiae* protein. However, unlike for ScNob1, the His-tag was cleaved with 3C protease after the first step. All buffers contained glycerol.

A) Initially, nickel affinity was used as for ScNob1.

B) Heparin sepharose purification. Again, similar to results seen for ScNob1.

C) As with the other Nob1 protein, contaminants are still seen at the SEC step with likely RNA contamination coming out in the void volume (i.e. first peak).

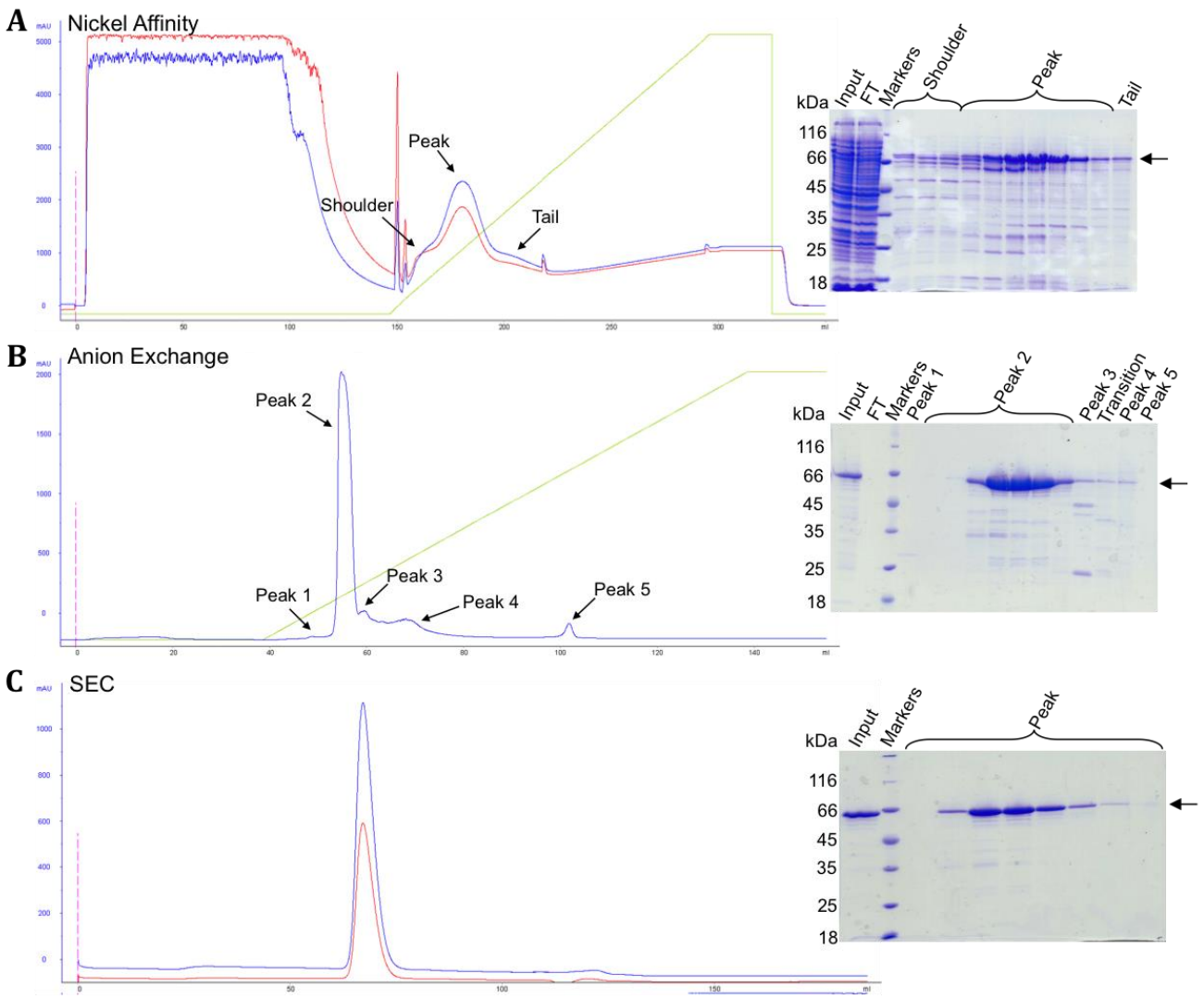


Figure 3.8: Optimised purification of CtNob1.

All the labelling and colour schemes are the same as on previous figures. No glycerol was used in the buffers and the second step was changed.

A) The nickel affinity column elution profile is similar to before. The first spikes in absorbance seen after the start of the gradient were artefacts due to pump malfunction during the run. The gel shows that, at this stage, more contamination is present without glycerol in the buffers.

B) A Mono Q anion exchange column was used to replace the heparin sepharose step. This gave very good separation with many contaminants coming off the column later in the gradient. The gradient ranged from 50 to 1000 mM salt.

C) The SEC shows only one peak, which corresponds to CtNob1 and its degradation products.

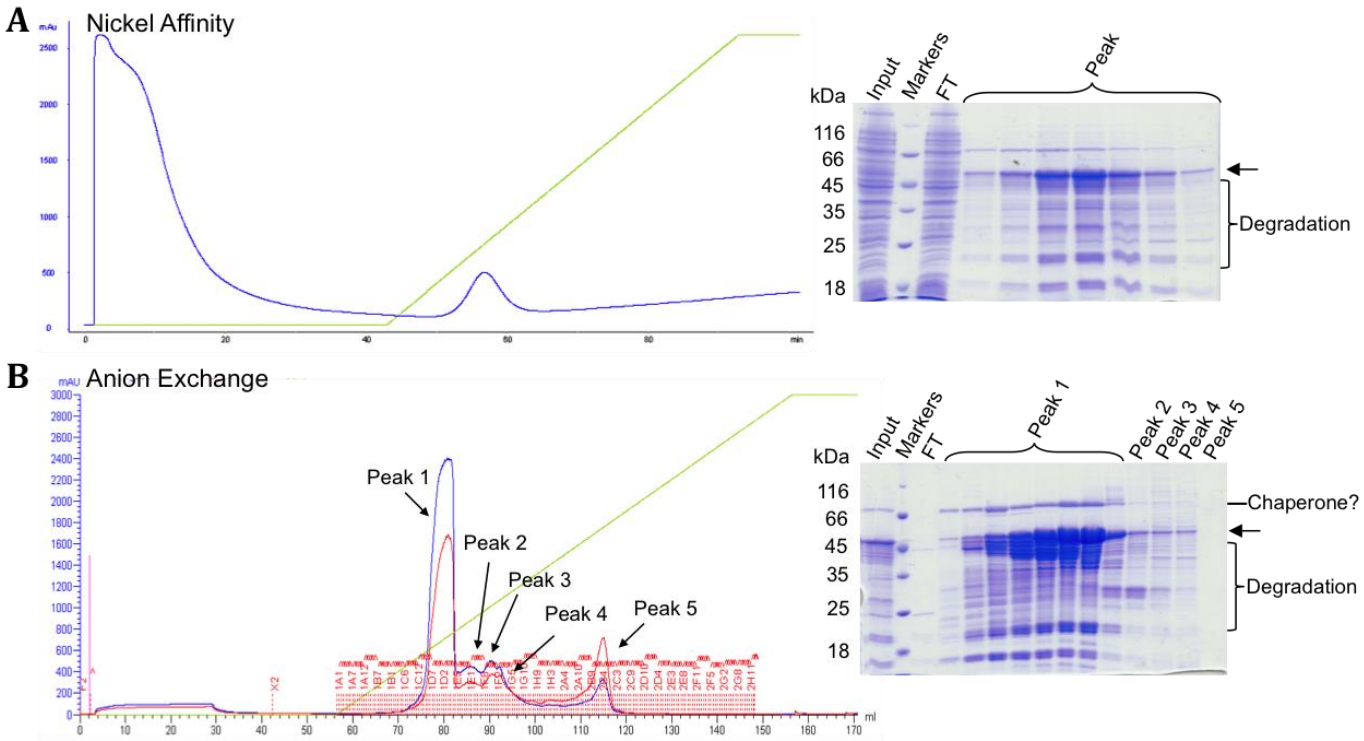


Figure 3.9: Purification of N-terminally tagged HsNob1.
The labelling is the same as before.
A) The nickel affinity elution profile shows a single peak. However, when the fractions are examined on the gel, many degradation products can be seen.
B) Elution from the anion exchange column (over a 100-1000 mM salt gradient). Due to technical issues during the run, fraction numbers are overlaid over the chromatogram in red. Multiple peaks were seen, most of them containing HsNob1, as shown on the gel. Also, a potential chaperone was pulled out and none of the degradation products were removed.

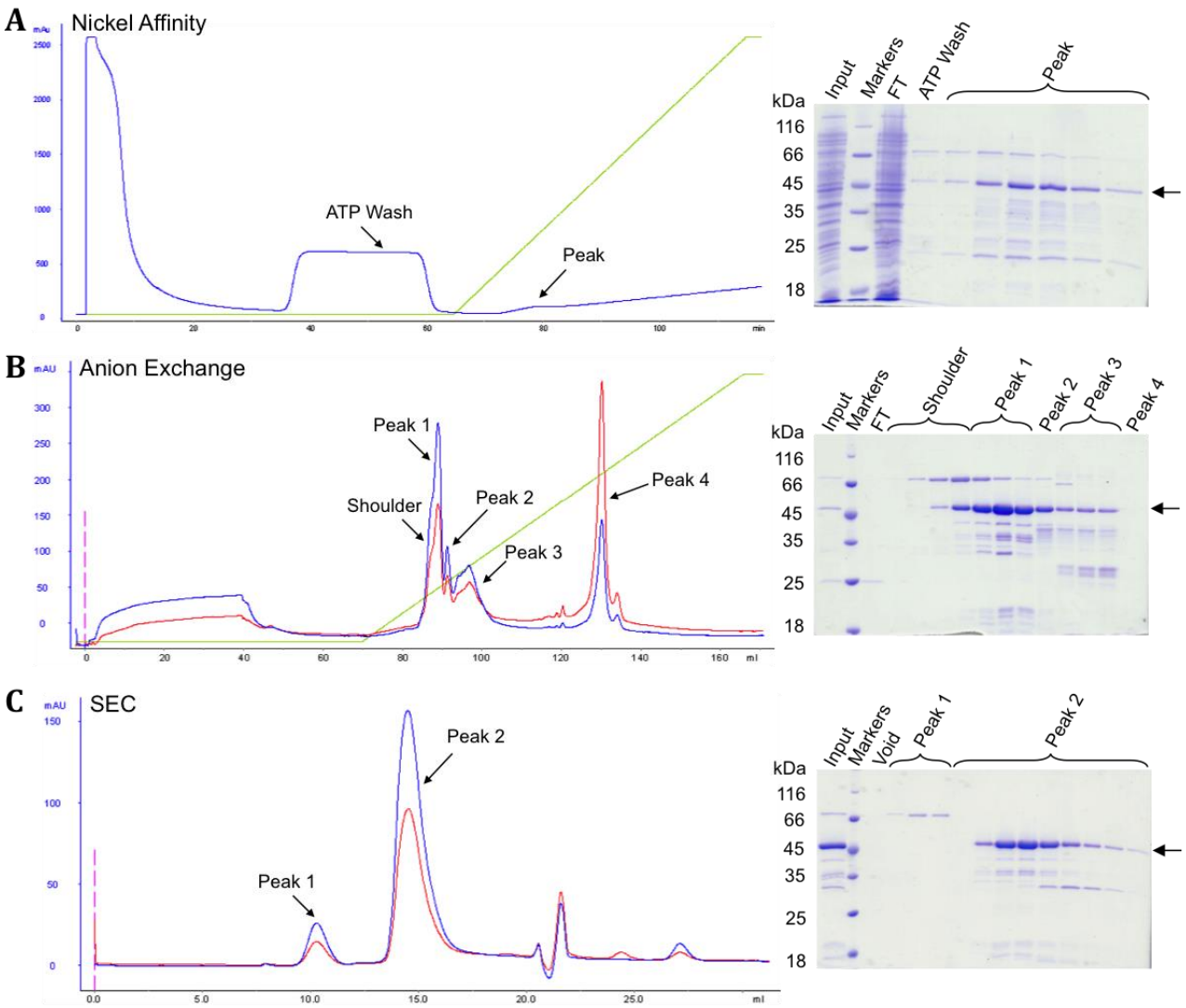


Figure 3.10: Purification of C-terminally tagged HsNob1.

A) An ATP wash was included in the nickel affinity step to remove the potential chaperone. A small peak containing HsNob1 was eluted. The gel on the right shows a much cleaner product than obtained with the N-terminally tagged construct.

B) The anion exchange profile, again, shows multiple peaks most of which contain HsNob1. However, most of the protein of interest is in peak 1. Therefore, fractions containing this peak were pooled and used in SEC.

C) Results of the SEC. Two peaks could be observed. One corresponded to the potential chaperone and the other to HsNob1. Although it is not completely pure, it looks much cleaner than the other construct.

3.3.4 ScNob1 – D15N mutant

After the WT protein was purified, the catalytic mutant (ScNob1 D15N) was subjected to the same purification protocol. Some of the protein did bind to the nickel beads but the majority was left in the flowthrough, supporting the results from the expression test. The protein that did bind was then dialysed into 100 mM salt in preparation for the cation exchange column. However, unlike the WT ScNob1, the mutant precipitated and it could not be rescued with higher salt. Therefore, the concentration of salt during dialysis was changed to 150 mM, which prevented the precipitation.

The mutant ScNob1 was then separated on a cation exchange column followed by SEC. During SEC, the protein came out in the void volume suggesting that it was aggregated. Also, the absorbance at 260 nm was higher than at 280 nm showing that RNA contamination was present (**Figure 3.11**).

In order to prevent the aggregation, the purification was performed again but with a sonication step added following cell lysis. As seen in **Figure 3.12**, this was successful in giving some non-aggregated protein. However, only 90 µg of the mutant ScNob1 were obtained from 4 litres of culture.

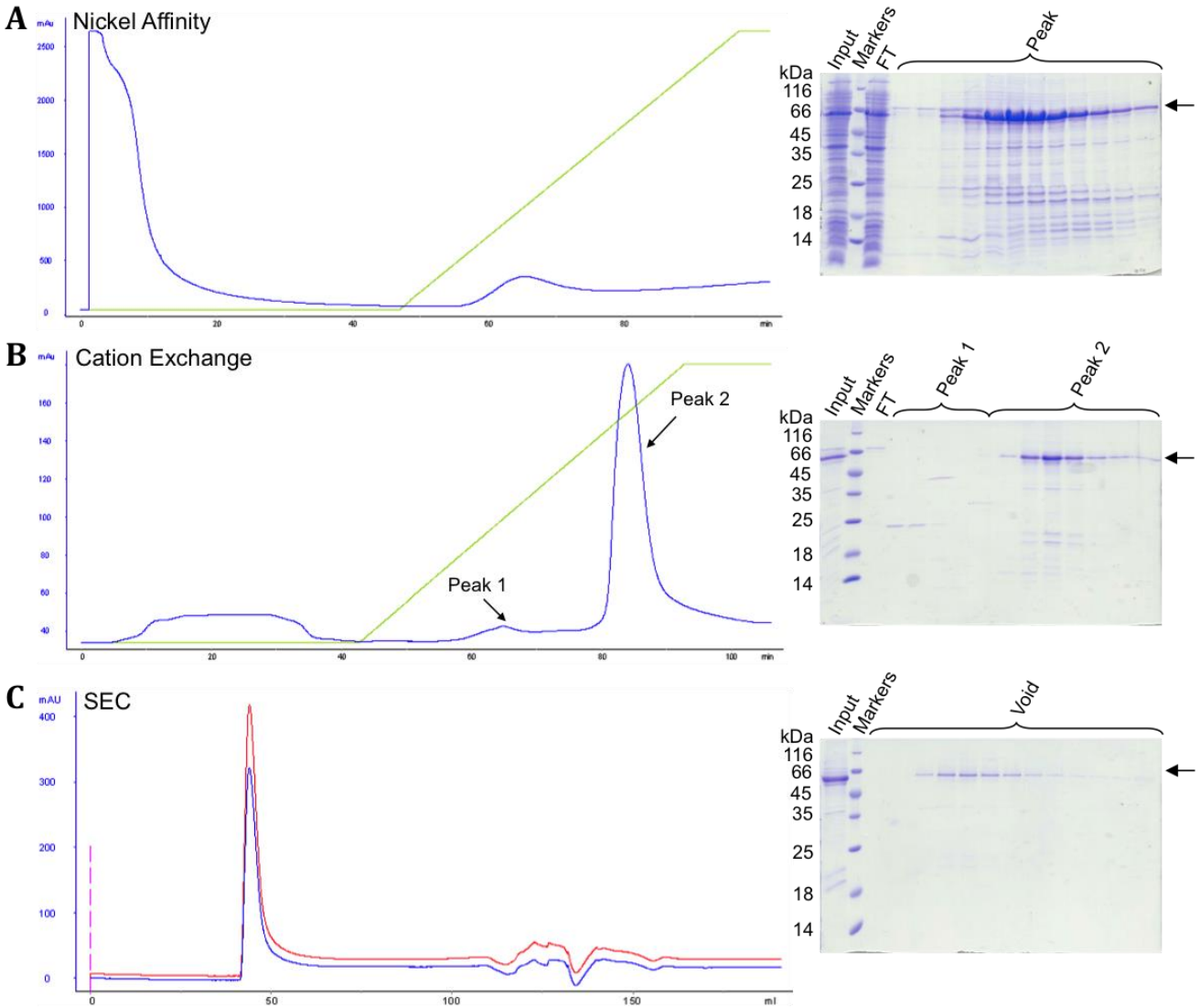


Figure 3.11: Initial purification of the ScNob1 D15N mutant.

A) The elution from the nickel column shows a single peak. When the fractions were separated on SDS-PAGE, a protein of the correct size could be seen. However, the majority still remained in the flowthrough (FT). Also, many contaminants were still present after this step.

B) Cation exchange was used as the second step. All of the protein bound to the column and was eluted late in the 150-1000 mM salt gradient. Some of the contaminants were removed using this step.

C) Again, SEC was used as the final purification step. However, all of the protein was eluted in the void (i.e. FT). Also, the absorbance at 260 nm (red) was relatively high.

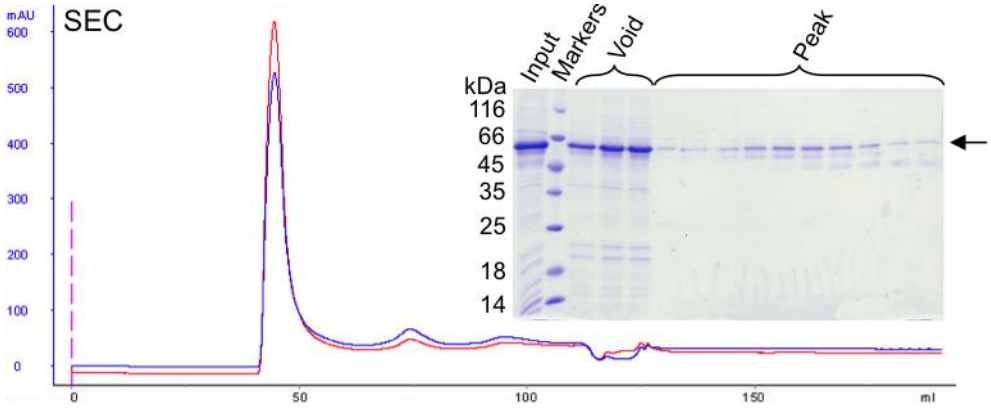


Figure 3.12: SEC results for the ScNob1 D15N mutant when sonication is used after cell lysis. The first two steps were identical and no difference was seen in the chromatograms or on the gels. The SEC chromatogram still shows a lot of protein in the FT and this was confirmed on the gel. However, a small portion of it elutes at the predicted volume. It still contains some contaminants or degradation products.

3.3.5 ScNob1 – Loop-out mutants

Initially, only two of the loop-out mutants were purified. ScNob1ΔL4 (Δ141-176), the mutant with the largest deletion, was purified first using the protocol established for the WT protein, with the exception that anion exchange was used instead of cation exchange. This change was necessary because the flexible loop that was removed by the mutation contained multiple positively charged residues. Thus, the new construct had a significantly reduced pI (6.1). The mutant did not bind to the anion exchange column as seen in **Figure 3.13**. However, the contaminants did bind, giving relatively pure protein by a subtractive purification step. After the SEC step, 2 mg of protein were obtained from 2 litres of culture.

ScNob1ΔL3 (Δ151-176) was then purified in the same manner (pI of 6.5). Again, relatively pure protein was obtained (**Figure 3.14**). In the end, 0.4 mg of protein were obtained from 2 litres of culture. The SEC profile showed some low molecular weight contaminants, likely nucleic acids.

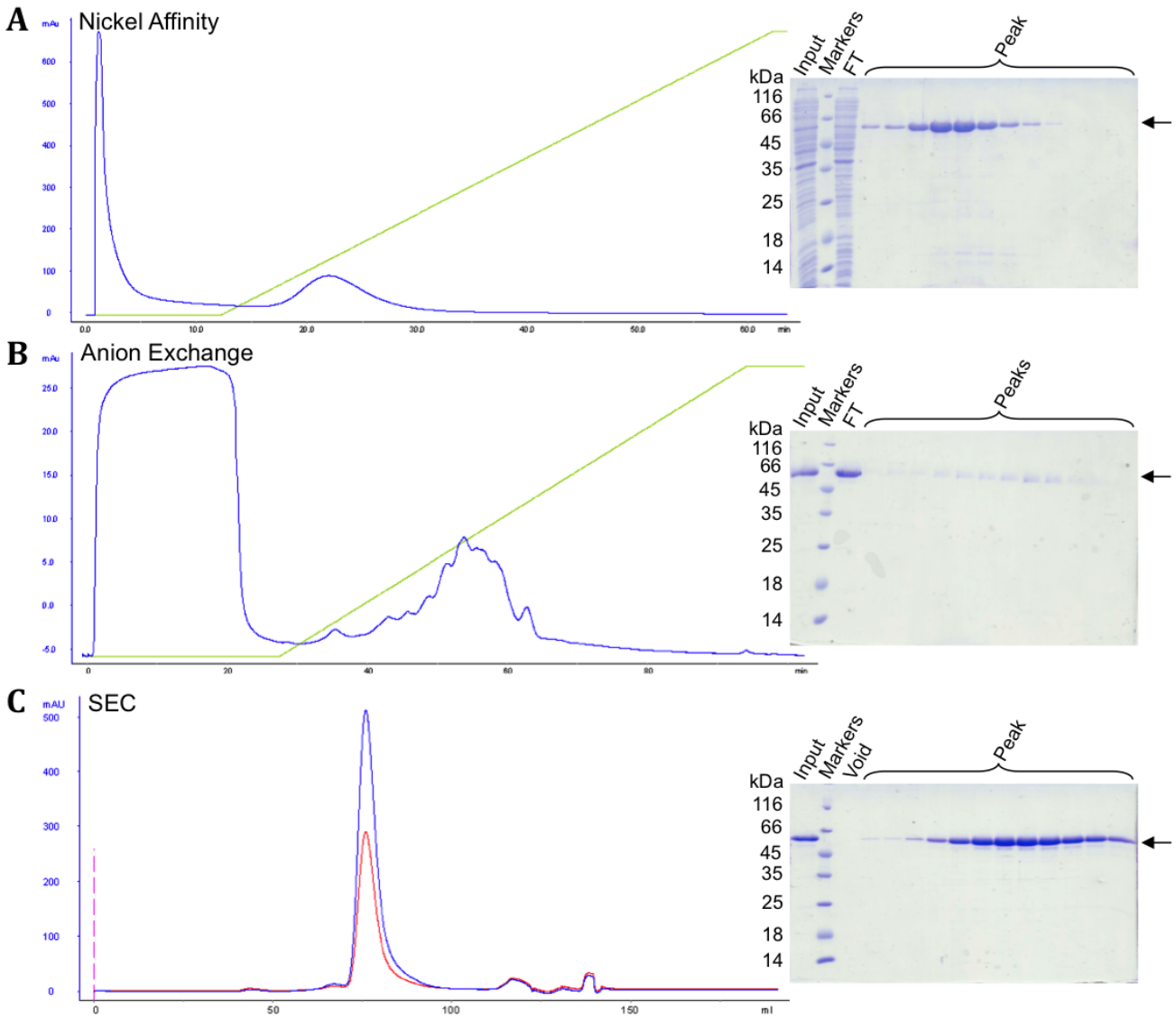


Figure 3.13: Purification results for the ScNob1 ΔL4 mutant.

A) A single peak is eluted from the nickel affinity column. This peak corresponds to the right protein and shows a relatively clean elution on the gel.

B) The sample was then injected onto an anion exchange column to which it did not bind. It could be that the starting salt (150mM) concentration was too high. The gel confirms that the majority of the protein remained in the FT. This was concentrated for SEC.

C) A single peak was obtained in the SEC that corresponded to the ScNob1Δ4 mutant on the gel. The protein appears to be relatively pure.

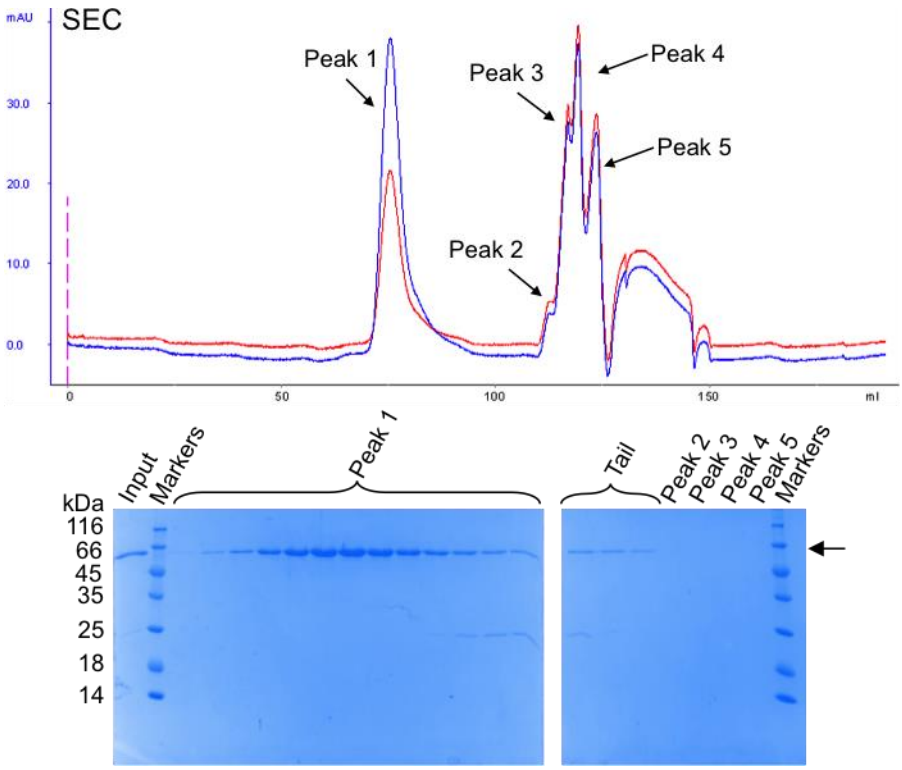


Figure 3.14: SEC profile for the ScNob1ΔL3 mutant.
The first two steps looked the same as for the ΔL4 mutant. In the last step, the protein was eluted in a single peak. However, there was a contaminant present, which gave rise to a tail. Also, low molecular weight contaminants were seen (peaks 2-5). These came out towards the end of the column volume (120 ml). Nothing is seen in these fractions on the Coomassie stained gel. The absorbance at 260 nm is slightly higher than at 280 nm and, thus, it is possible that the contaminants are RNA.

Finally, with the help of a summer intern (Michal Merdas), the other loop-out mutants were also purified using the same protocol. The protein parameters and yields from 2 litres of culture are listed in **Table 3.3**. All of the constructs behaved in the same manner throughout the purification. ScNob1ΔL1 showed some aggregation in the SEC step performed by Michal Merdas (data not shown). However, a peak of the right size was still seen.

Table 3.3: Summary of loop-out protein parameters and the yield obtained from 2 litres of bacterial culture.

Construct	Molecular Weight (kDa)	Isoelectric point	Yield from 2L (mg)
Nob1ΔL1 (Δ151-159)	52.0	6.8	2.4
Nob1ΔL2 (Δ151-169)	50.9	6.3	1.2
Nob1ΔL3 (Δ151-176)	50.2	6.5	0.4
Nob1ΔL4 (Δ141-176)	49.1	6.7	2.0
Nob1ΔL5 (Δ141-159)	50.8	7.0	2.3

3.4 The ScNob1 Protein is Unlikely to be Monomeric

ScNob1 was purified and eluted as a single peak in SEC. However, the size appeared to be larger than predicted for the monomer (53 kDa). **Figure 3.15A** shows a comparison of the ScNob1 SEC profile with some known size standards. However, as SEC depends on the shape of the molecule as well as the size, the oligomeric state of ScNob1 cannot be obtained from this result.

Thus, a sample of ScNob1 was sent to Dr. Andrew Leech at the University of York for size exclusion chromatography coupled with multi-angle light scattering (SEC-MALS) analysis, which can give absolute molecular mass. The resulting calculations showed a size of about 120 kDa. However, this was an overestimate, as there seemed to be a mixture of high and low molecular weight material (**Figure 3.16**). Overall, it appears that ScNob1 might form a dimer in solution.

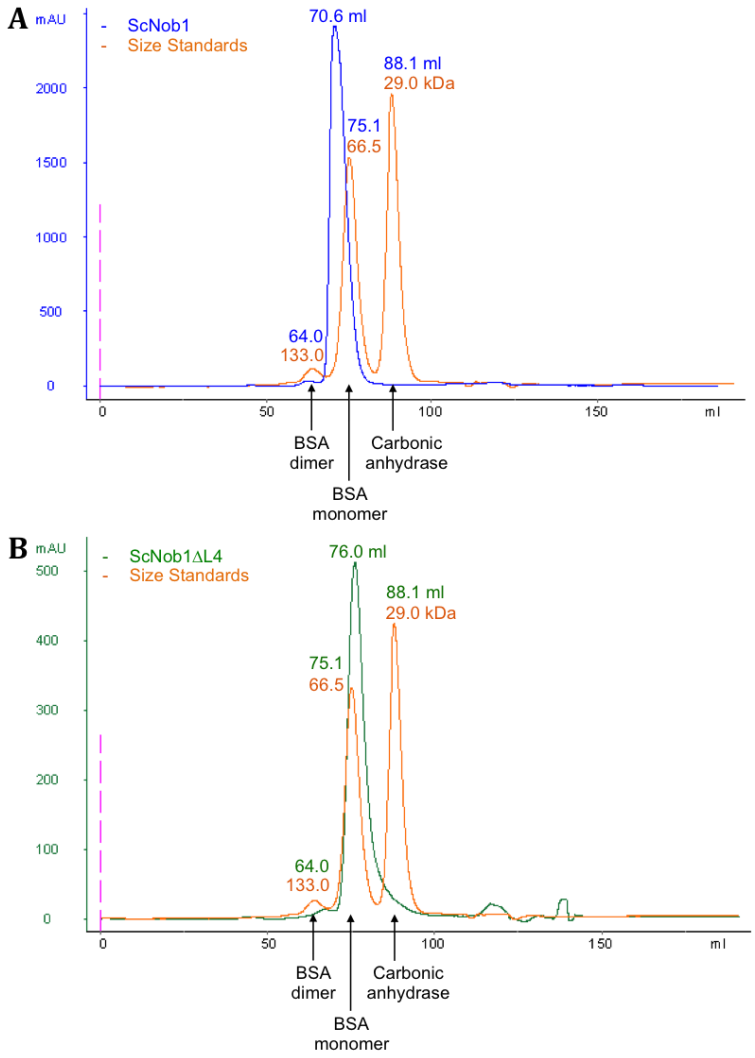


Figure 3.15: ScNob1 and ScNob1ΔL4 SEC results compared to size standards. The protein standards are labelled under the chromatogram. Their sizes (in kDa) are shown in orange next to each peak with the corresponding elution volumes shown above these in blue or green. **A)** Comparison of ScNob1 with the size standards. ScNob1 appears larger than the BSA monomer but smaller than the dimer. As the proteins get larger, the space between peaks gets smaller. **B)** SEC results for ScNob1ΔL4 compared to the standards. The protein is now smaller than the BSA monomer (66.5 kDa). Thus, it is either significantly smaller than the WT or it has changed its shape.

The loop-out version (ΔL4) of ScNob1, however, shows a relatively large shift to a later elution volume in the SEC (**Figure 3.15B**) suggesting that the size had significantly changed. The deletion in ΔL4 should correspond to a decrease in size of about 3.3 kDa. However, the difference seen in SEC suggests a larger change, possibly a monomerisation. The elution volume of the ΔL3 mutant was almost the same as for ΔL4 (less than 0.6 ml difference), suggesting that the decrease in size between the mutants and the WT protein (5.4 ml to ΔL4) is significant. The ΔL3 mutant should be about 1 kDa larger than ΔL4 and 2.2 kDa smaller than the WT protein.

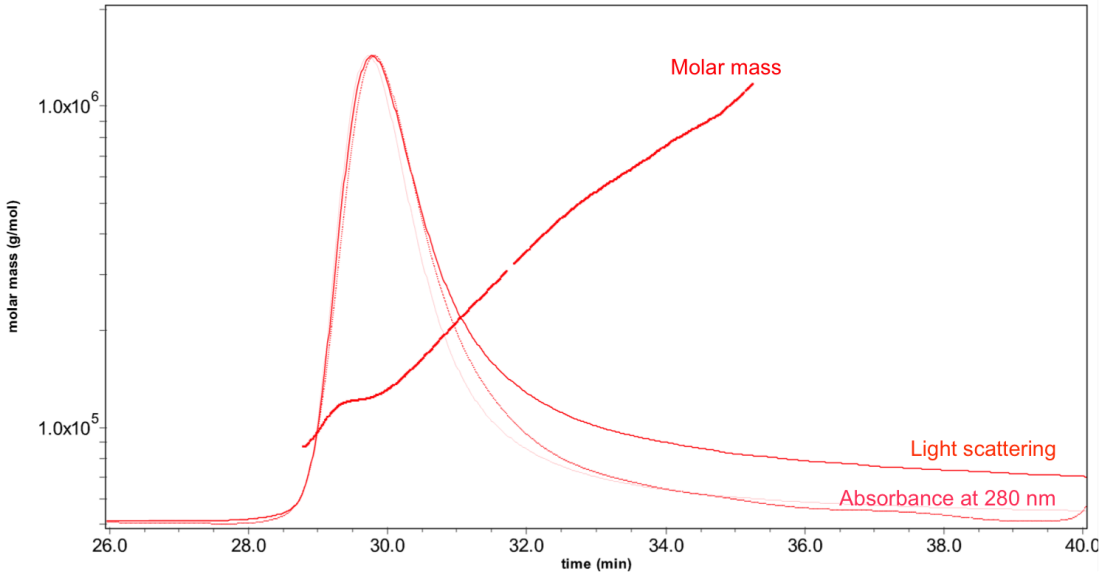


Figure 3.16: ScNob1 molar mass plot obtained from SEC-MALS analysis.
The plot shows the calculated molar mass as the overlaid dark red line and the SEC peak. The absorbance at 280 nm (top red trace) and the amount of light scattering (bottom red trace) is shown as the protein is eluted from the SEC column (x-axis shows the elution time). The y-axis displayed is associated with the molar mass line and is shown on an exponential scale. The y-axes for the absorbance and light scattering are not shown for simplicity. The flat portion of the molar mass line at about 29.5 min corresponds to about 120 kDa. However, the line then goes up exponentially suggesting the presence of high molecular weight particles along with the smaller molecules. This is supported by the presence of a tail following the peak in the chromatogram.

3.5 Determination of Protein Stability by Thermal Denaturation

3.5.1 ScNob1

The stability of the purified ScNob1 was tested using thermal denaturation in different buffer conditions. In this assay, Sypro Orange is used as a dye that increases in fluorescence when it is in a hydrophobic environment. Thus, when a protein denatures at higher temperatures and exposes its hydrophobic core, an increase in fluorescence is seen. This allows the temperature at which the protein is denatured (melting temperature or T_m) to be obtained, giving a measure of its stability in different conditions.

Initially, it was found that the T_m of the protein decreased from 48.5˚C to 45.7˚C with increasing concentration. Also, the melting temperature decreased with lower pH or lower glycerol concentrations. However, salt and zinc concentrations had no effect on protein stability (**Figure 3.17**).

3.5.2 CtNob1

As expected, the *C. thermophilum* protein was more stable than the yeast protein (T_m = 61.0˚C vs. 48.0˚C). Protein concentration did not have as significant an effect on stability as for ScNob1. Increasing salt concentration changed the stability by a modest amount (60.7˚C to 61.7˚C over 50 to 250 mM NaCl). Also, CtNob1 showed a decreased melting temperature in the presence of 10 μM zinc (61.0˚C went down to 60.2 ˚C). Like the yeast protein, decreasing pH and glycerol also decreased the melting temperature of CtNob1 (**Figure 3.18**).

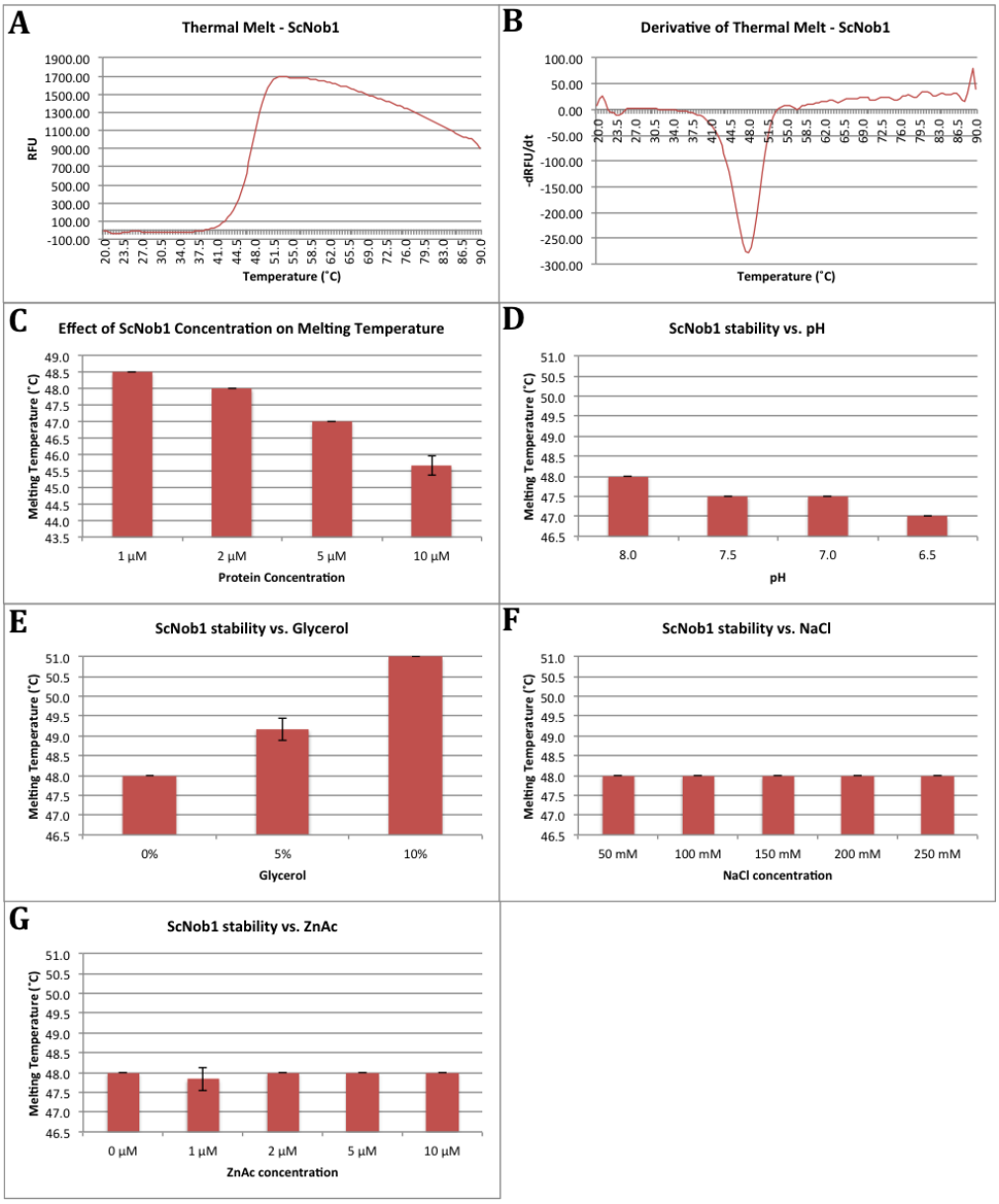


Figure 3.17: Thermal denaturation results for ScNob1.

The experiments were performed in triplicate. The error bars represent the standard deviations.

A) An example of the raw thermal melt obtained. The increase in relative fluorescence units (RFU) was recorded over changes in temperature. SYPRO Orange fluorescence increases as it comes in contact with hydrophobic residues. Therefore, the increase in RFU corresponds to protein denaturation. For ScNob1, a single melt curve is seen.

B) An example derivative of the thermal melt. The bottom of the peak corresponds to the melting temperature (T_m). This value was taken for each of the experiments and averaged over the three repeats to obtain the summarised results.

C) Summary of the effect of protein concentration on protein stability. Higher ScNob1 concentrations led to a lower melting temperature suggesting decreased stability.

D) 2 µM protein was used for all the following experiments. Summary of the effect of pH on the melting temperature. The protein appears to be more stable at higher pH.

E) Effects of glycerol concentration on ScNob1 stability. The T_m increases with glycerol.

F) NaCl concentration has no effect on protein stability.

G) Zinc has no effect on ScNob1 stability.

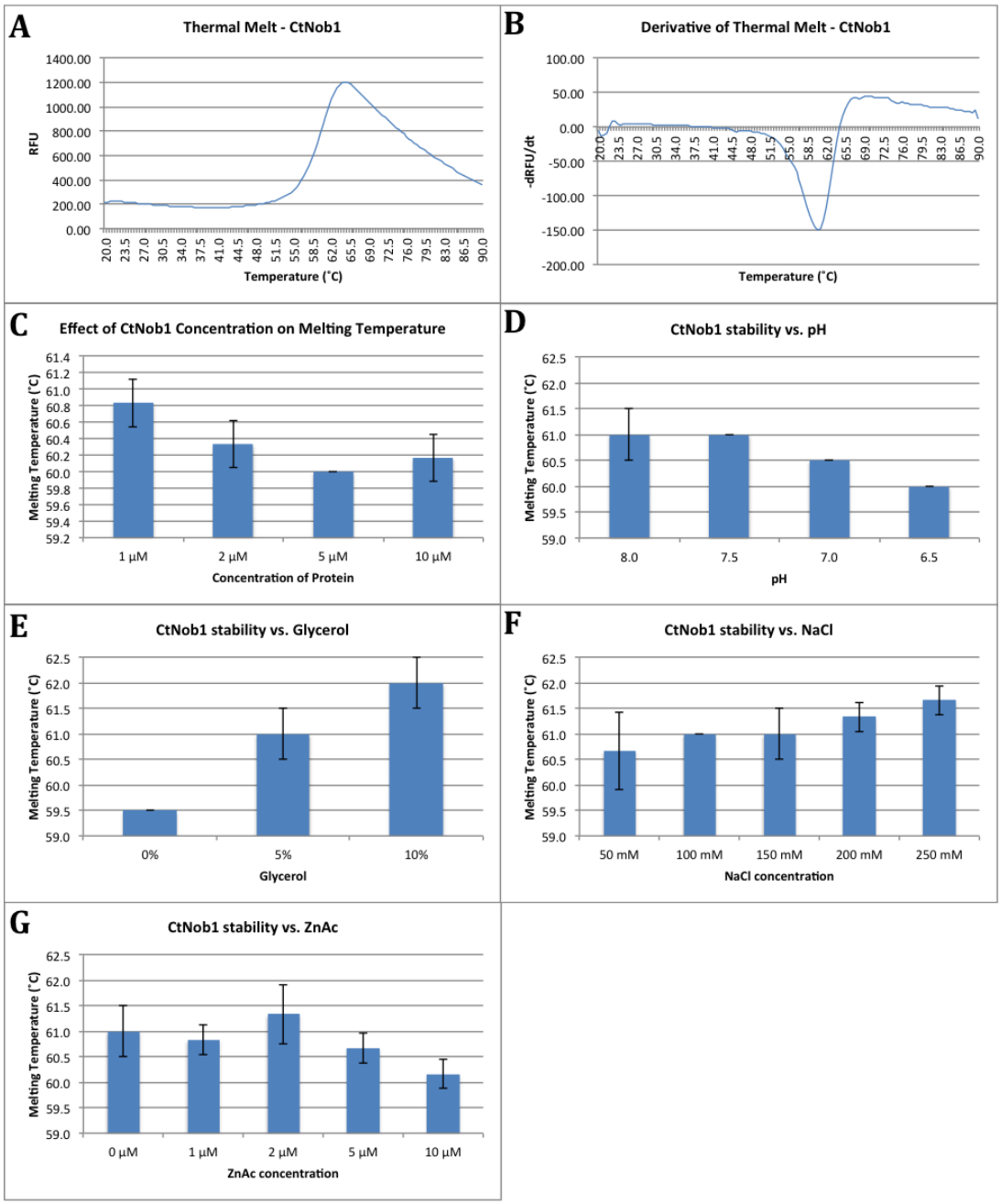


Figure 3.18: Thermal denaturation results for CtNob1.

The experiments were performed in triplicate. The error bars represent the standard deviations.

A) An example of the raw thermal melt obtained. As for ScNob1, a single melt curve is seen for CtNob1.

B) An example derivative of the thermal melt. The T_m is higher than that of ScNob1.

C) Summary of the effect of protein concentration on protein stability. As for ScNob1 concentrations led to a lower melting temperature but the effect is less pronounced.

D) 2 µM protein was used for all the following experiments. Summary of the effect of pH on the melting temperature. This protein also appears to be more stable at higher pH.

E) Effects of glycerol concentration on CtNob1 stability. The T_m again increases with glycerol.

F) NaCl concentration slightly increases CtNob1 stability (unlike for ScNob1).

G) Zinc has no clear effect on CtNob1 stability. However, a lower T_m is observed at 10 µM zinc concentration.

3.6 Identification of Stable Domains by Limited Proteolysis

3.6.1 CtNob1

CtNob1 was digested with a series of dilutions of different proteases (**Figure 3.19**) in order to determine if any stable domains are present. The low concentrations of proteases should allow cleavage only in exposed flexible regions, whereas stable folded domains would be protected. These would then be identified as abundant bands on Coomassie-stained SDS-PAGE gels.

Chymotrypsin gave one relatively stable fragment. MALDI-ToF analysis of the band identified peptides that primarily map to the N-terminus, suggesting a C-terminal truncation. However, due to time constraints, this was not explored any further. The same experiment had previously been performed with ScNob1 by Dr. Atlanta Cook without any promising results and was, therefore, not repeated.

3.7 Crystallisation Trials

Crystallization trials of ScNob1 and CtNob1 were performed using several commercial screens in 96-well plates (Magic I, Magic II, Hampton Index, Morpheus, JCSG+, and Structure screen 1 and 2). However, none of them yielded any promising crystal hits. This is likely due to the proteins' flexibility.

All of the loop-out mutants were also tested (using at least the Hampton Index and Magic I screens for each) with the help of Michal Merdas. The initial results looked more promising than for the full-length constructs. Over half of the wells for the mutants showed granular precipitation, which suggests that nucleation events could be happening. About a week after screening, crystals were observed in one condition for ScNob1ΔL2 and ΔL3. These are shown in **Figure 3.20**. The condition contained BIS-TRIS at pH 6.5 and 5.5 respectively and 25% (w/v) PEG 3350.

The crystals seen were very small and needle shaped. These need to be further optimised in the future. Due to time constraints and their size, they have not yet been tested for diffraction.

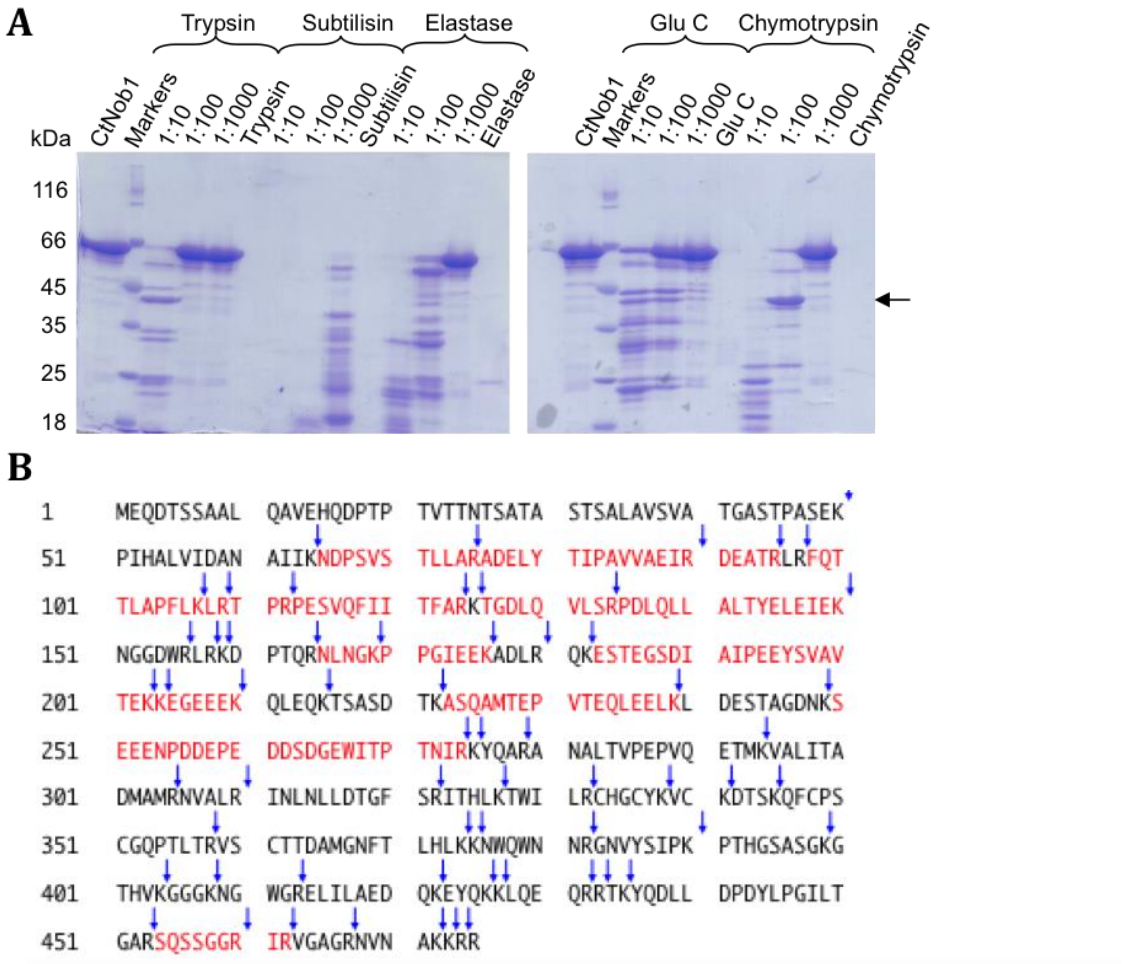


Figure 3.19: Limited proteolysis and mass spectrometry analysis of CtNob1.

A) Results of limited proteolysis performed on CtNob1. The enzymes used are shown above. Three dilutions per enzyme (of 1 mg/ml stocks) were used. Limited proteolysis allowed mapping of stable regions. The band of interest in the 1:100 chymotrypsin lane is marked with an arrow.

B) Mass spectrometry analysis of the band of interest. The band was excised from the gel and subjected to trypsin digestion. The predicted cleavage sites are marked with blue arrows. Then, the resulting peptides were analysed using MALDI-ToF mass spectrometry. The identified fragment is likely a C-terminal truncation. Most of the N-terminal portion is covered by the identified peptides (shown in red). Only one C-terminal peptide is seen suggesting that this part of the protein is underrepresented in the sample, even though some of the predicted fragments would be long enough to be seen. The experiment can be potentially scaled up in order to obtain a large amount of the truncated protein.

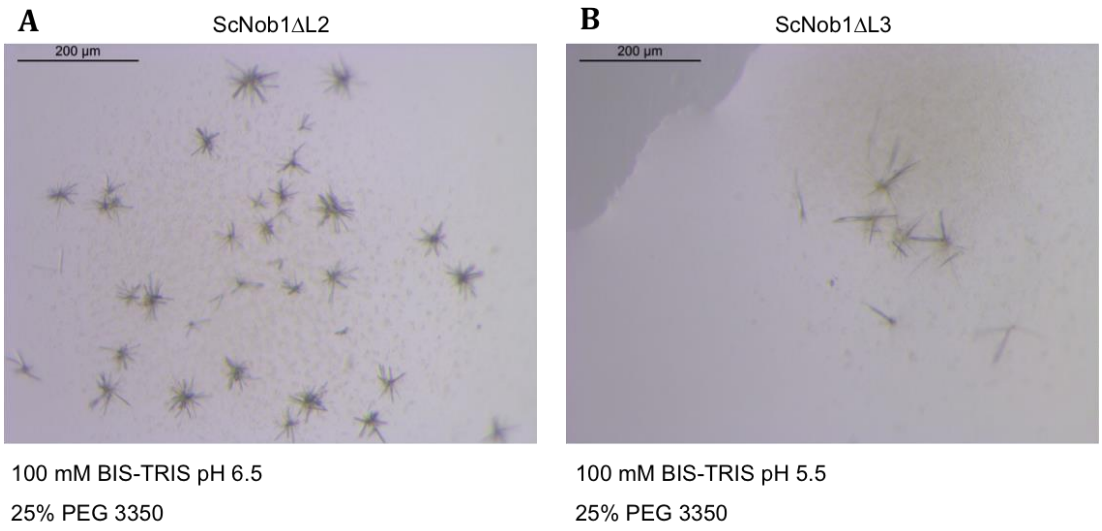


Figure 3.20: Initial crystal hits obtained with ScNob1ΔL2 and ΔL3.
A) The small needles obtained using ScNob1ΔL2 in the Hampton Index screen. The crystals are very small and attached to each other.
B) Crystals obtained with ScNob1ΔL3 in the same screen but at a different pH. The needles are larger but appear less straight.

3.8 Discussion

3.8.1 Protein expression and purification

Nob1 from three organisms was successfully expressed and purified. However, the level of success varied between the different constructs. The yeast Nob1 was over-expressed in SE1 cells from the pSL6 vector. However, when the same protein was re-cloned into the pET28α plasmid, the expression level dropped. A similar expression level to the pSL6 construct was observed in the expression tests using the RIPL cell line. However, this still equated to significantly less protein per unit biomass.

Successful purification protocols were established for both constructs. DTT was replaced by β-mercaptoethanol in all Nob1 purifications, as it has been shown to be able to chelate zinc in the past (Krezel et al, 2001). The protocols were optimised so that the SEC acted as final quality control. After the cation exchange, all the visible contaminants were removed. However, some low molecular weight bands could still be seen on the gel. The fact that they all came out in a single peak in SEC

suggests that they were degradation products. Their amount was minimised in later preparations by always keeping the sample at 4˚C during purification and performing the first two steps as quickly as possible.

Human Nob1 was clearly over-expressed when it was cloned into the pEC-KHC vector. However, there were multiple degradation products eluted after the first step and purification led to a mixture of differently sized fragments that could not be separated. Therefore, the protein was cloned into a pET28α plasmid with a C-terminal His-tag. This helped to decrease the amount of degradation products. However, while the protein was still enriched on nickel beads, it was less well expressed than the other construct.

Nevertheless, the C-terminally tagged HsNob1 was purified. Due to the low level of expression and the presence of multiple contaminants, a very small amount of protein (30 µg from 4 litres of culture) was obtained. Thus, this protein was not used for any assays or for crystallisation.

Nob1 from the thermophilic fungus, *C. thermophilum*, was ordered as a codon optimised construct. Thus, it was used directly in expression trials where it was shown to be over-expressed in all the strains used. The purification protocol used was the same as for ScNob1 and it was optimised in the same fashion. As with the yeast protein, some degradation was seen but it was reduced over time by performing the purification fast and keeping the protein on ice.

For human and yeast Nob1, the catalytic mutants were expressed (D10N and D15N respectively). Surprisingly, the N-terminal mutation decreased the ability of the C-terminal His-tag on ScNob1 to bind to nickel beads. The D15N mutant was clearly over-expressed and soluble, but it was not enriched on the beads. During purification trials, after the nickel affinity step, most of the protein was still in the flowthrough. It was later shown in the SEC that the mutant formed an aggregate. Thus, it could be that the C-terminal His-tag was sequestered in the aggregate and could, therefore, not bind to the resin.

With the human D10N mutant, the same trend was seen with the N-terminally tagged protein. However, as the mutation is on the N-terminus, there are potential changes in conformation that could not have been predicted and perhaps

sequestered the His-tag. Purification of this construct was not undertaken due to the problem with degradation products mentioned earlier.

The C-terminally His-tagged D10N mutant showed an ability to bind to nickel beads that was comparable to the WT HsNob1 protein. However, the over-expression was low as for the WT. Therefore, no further attempts were made to purify this construct either.

Finally, five yeast loop-out mutants were tested. All five mutants showed more over-expression in the pET28α vector than the WT ScNob1. All of these were purified. The first one used, ΔL4, had the longest truncation. A single peak corresponding to the relevant protein was seen for each mutant in SEC. This suggests that the mutant proteins still folded properly. The elution profile for the ΔL3 mutant showed a tail and substantial low molecular weight contaminants (likely RNA). However, the SEC peak still contained relatively pure protein. ΔL1 showed some aggregation in the void volume. However, a peak for the non-aggregated protein was also seen and contained the pure product. The other two mutants behaved well, giving single peaks in SEC containing pure protein.

3.8.2 Oligomerisation state of Nob1

Initially, when the full-length Nob1 was purified, its size (predicted on the elution volume from the SEC column) appeared to be larger than expected. However, the separation in SEC depends on the shape of the molecule and its interactions with the matrix as well as its size (Wen et al, 1996). Therefore, it is possible that the Nob1 protein is a monomer, but it is extended in one direction giving the appearance of a larger molecule. The archaeal structure showed that the two domains move with respect to each other due to the flexible linker. Thus, at any time, they might be relatively far apart from each other creating an elongated molecule.

SEC-MALS was performed in order to obtain a definite size of the particles. This method relies on light scattering immediately following the SEC. The scattering of the laser is measured at different angles and is dependent on the properties of the particles in the solution. Unlike with just SEC, the absolute molecular weight can then be calculated from the scattering behaviour (Oliva et al, 2001).

Results from the SEC-MALS performed at the University of York were inconclusive as well, as the sample appeared heterogeneous. The presence of the tail following the peak suggests that there are some interactions between the protein and the matrix. The experiment was performed at room temperature, which could also have caused additional changes in the protein-protein and protein-matrix interactions compared to the purification performed at 4˚C. Nevertheless, it appears as though the particle is around 120 kDa, although this is likely an overestimation. The results would be consistent with ScNob1 being a dimer in solution.

In the future, if possible, SEC-MALS should be repeated with fresh protein immediately from purification. Also, to complement the method, the same sample (once retrieved from SEC-MALS) should be tested with dynamic light scattering (DLS). This technique measures the dynamic properties of the particles in the sample, such as the hydrodynamic radius. Like with MALS, scattered light is measured. However, the information is obtained from the effect of the neighbouring particles and Brownian motion of the protein on the scattered light. This would give an idea about whether the protein is monodispersed and about how many other species are present in the sample (Berne & Pecora, 2000). This would then allow clarification of ambiguous SEC-MALS results where more than one species appears to exist.

The thermal denaturation experiments showed that ScNob1 stability is concentration dependent. For all of the SEC experiments, the protein concentrations were relatively high (about 38 μM). Therefore, it is possible that this caused some protein to be improperly folded giving an asymmetric peak in SEC.

Additionally, the SDS-PAGE gels from SEC showed the presence of smaller species, which are likely to be degradation products. This would be supported by the fact that a single elution peak is seen suggesting that the smaller species are associated with each other making a complex of a similar size/shape as the full-length protein. For example, if the proteolysis occurred in the flexible linker, the two functional domains could stick together through protein-protein interactions in SEC but be separated in denaturing gels. These types of species could also cause the tail seen in SEC. The identity of the fragments could be confirmed by mass spectrometry analysis of the bands.

Conversely, the two loop-out mutants purified showed elution volumes in SEC that would be consistent with the protein being monomeric. Thus, it is possible that the flexible loop region is responsible for dimerization and removing it prevents this from happening. This is unlikely as this part is not well conserved and it does not appear to be folded. Nevertheless, SEC-MALS will have to be performed on these mutant proteins in order to obtain the absolute size.

3.8.3 Potential for crystallisation

All of the purified constructs were used for crystallisation screening where possible. The full-length proteins did not give any hits. These proteins have a long flexible region between the two functional domains. In archaea, this region was shown to be very flexible in NMR (Veith et al, 2012). Thus, it is likely that the protein is too dynamic for crystallisation.

To negate this issue, the loop-out mutants were created. If the two domains interact with each other, these mutations would help by removing the flexible region (**Figure 3.21A**). This arrangement is supported by the fact that when the domains were cloned separately (by Iva Tchasovnikarova), they were found to be insoluble. However, it is also possible that the two domains do not form any interactions and rotate independently of each other. In this situation, the deletion of the linker would only bring the domains closer together without making the protein less flexible (**Figure 3.21B**).

All of the mutant proteins were purified and two of them gave a hit in the Hampton Index screen. Also, all of the constructs behaved well in purification, implying that the folding of the domains was not significantly changed. No aggregates were present (apart from ΔL1) and the crystallisation screens showed the presence of globular precipitation. In the future, the initial crystals obtained should be optimised and grown on a larger scale.

Once they are large enough, the crystals can be tested for diffraction at the Diamond Light Source synchrotron (Oxford, UK). It is possible that the species forming are salt crystals but this is unlikely, as the crystallisation conditions did not contain additional salts. Nevertheless, this cannot be ruled out at this stage and further testing needs to be undertaken. If the initial crystals do not diffract, the

species can be identified by dissolving the crystals, separating them on SDS-PAGE, and silver-staining the gel.

The CtNob1 protein showed a promising stable fragment in limited proteolysis. Mass spectrometry analysis of the protein fragment showed large coverage of the N-terminus and only a single peptide at the C-terminus. Thus, this band could represent a C-terminal truncation. The last N-terminal peptide identified ends at residue 274. The flexible loop spans residues 163-321. Therefore, it is possible that the proteolysis occurs within the loop.

Large scale limited proteolysis coupled with SEC purification would lead to enrichment of this fragment and electrospray mass spectrometry could be used to obtain a total mass in order to identify the cleavage site. As it appears to be proteolytically stable, it might represent a single domain that could prove easier to crystallize.

Previously, Iva Tchasovnikarova has tried to express the individual domains (PIN and zinc ribbon) separately. However, they were found to be insoluble. Using the limited proteolysis, a better map for the domains could be obtained and soluble fragments might be able to be cloned. Due to time constrains, this has not been attempted during this project but would be useful in the future.

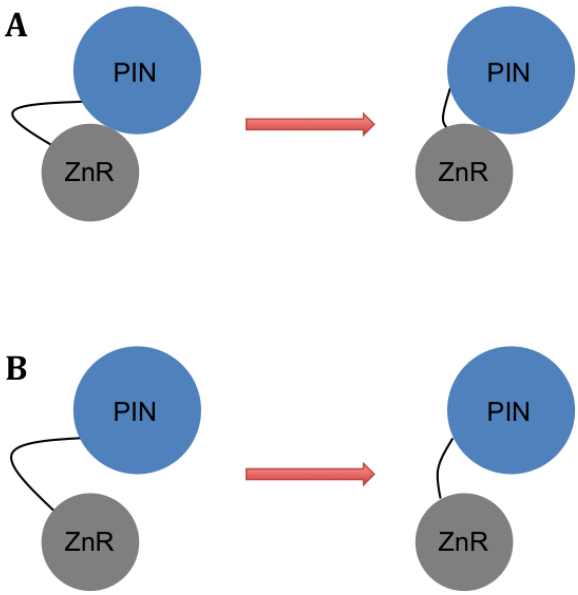


Figure 3.21: The two possibilities for domain organisation of Nob1.
The PIN (blue) and zinc ribbon (ZnR, grey) domains are shown to be connected by the flexible loop (black line).
A) If the two domains interact with each other, the deletion of the loop removes the flexible region of the protein particle
B) If the two domains do not interact, the deletion of the loop still allows them to rotate with respect to each other. Thus, the flexibility is still present.

3.9 Concluding remarks

Overall, the expression and purification of Nob1 were successful. The protein appeared to be folded and stable at room temperature. The oligomerisation state of each of the constructs remains in question. Crystallisation of the loop-out mutants has shown a promising hit with a lot of potential for optimisation in the future. The two domains could also be cloned separately in order to solve the individual structures. This was unsuccessful in the past but potential for optimisation still exists.

4 INVESTIGATING NOB1-RNA INTERACTIONS

4.1 Introduction

In order for Nob1 to perform its function, it needs to form the necessary contacts with RNA. This is especially important as its nuclease activity has to be specific to the D site (Pertschy et al, 2009). The interaction between Nob1 and RNA has been debated for a long time. Some studies have shown it binding near to the D site (Lamanna & Karbstein, 2009). However, initial cross-linking studies placed it at a more distant site (helix 40) that would require a conformational change to occur in order for the cleavage to take place (Granneman et al, 2010). Otherwise, the zinc ribbon domain could be binding to helix 40 while the PIN domain interacts with the D site.

More recent cross-linking experiments in growing cells, performed by the Tollervey lab, (unpublished results) support that Nob1 does interact with the D site. Also, a sequence upstream of the cleavage site, in the first internal transcribed spacer (ITS1), was identified using these conditions. Again, it is possible that all these interactions occur at different stages of pre-18S maturation and/or via different domains of the protein. Additionally, the predicted Dim2 binding sequence (GGAUCA) is proximal to the cleavage site, potentially obscuring or aiding Nob1 binding.

It is also unclear which domain of the Nob1 protein is more important for the RNA interactions. It has been shown that the PIN domain is required for the cleavage at site D (Fatica et al, 2004; Lamanna & Karbstein, 2009). The activity of this domain is stimulated by manganese cations (Lebaron et al, 2012). However, it is unclear if the domain is bound to the D site prior to the cleavage or if it only makes the necessary contacts immediately preceding cleavage.

Finally, although Nob1 has been shown to bind to the D site and some different sequences (Granneman et al, 2010; Lamanna & Karbstein, 2009), no consensus recognition sequence for Nob1 has been identified. Thus, it is unclear how it is properly placed on the RNA to perform its high fidelity function. In this chapter,

recombinant eukaryotic Nob1 is used to identify the potential short RNA sequences recognised by this protein.

4.2 Nob1 Does Not Interact with Helix 40 RNA *in vitro*

In order to investigate whether recombinant Nob1 can bind to the initially identified cross-linked site, short RNAs containing the helix 40 (H40) sequence were designed. Three RNAs of different lengths were used, as the exact recognition sequence is unknown. To transcribe the individual RNAs, the T7 polymerase was purified as described in the next section.

4.2.1 Expression and Purification of the T7 Polymerase

The T7 polymerase was used for the large and small scale *in vitro* transcription of different RNAs during this project. The protein expression was initially tested in M15 cells in two media: Superbroth and 2xTY. The expression was very good, as shown in **Figure 4.1**. Thus, 2 litres of culture were used to purify the polymerase according to a previously established protocol (Arts et al, 1997).

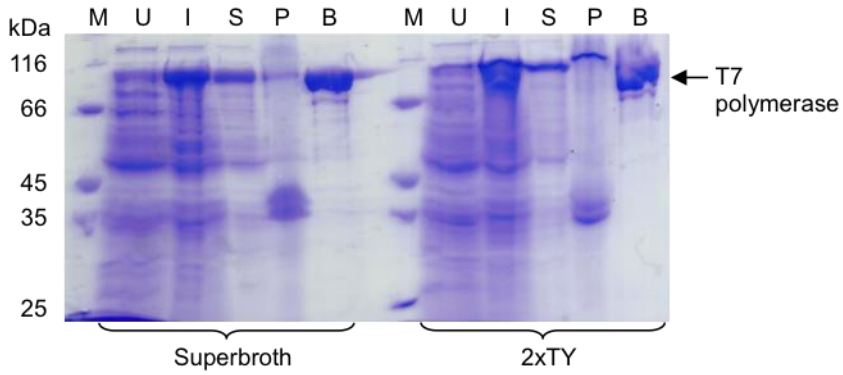


Figure 4.1: Expression trial of T7 polymerase in M15 cells in Superbroth and 2xTY media.
The different lanes correspond to the different fractions: uninduced (U), induced (I), soluble (S), insoluble/pellet (P), and the nickel bead-bound (B) fraction. The molecular weight markers (M) are also shown on each gel. The band corresponding to the correct size is indicated with an arrow. The protein is highly over-expressed in both media and enriched on the beads as expected. Superbroth was chosen as the medium for large-scale expression.

The first step of purification used affinity chromatography. The His-tagged protein was bound to cobalt beads and eluted with an increasing concentration of imidazole. As seen in **Figure 4.2**, a single large peak containing the polymerase was eluted. The protein was pooled and dialysed into a low salt buffer. Then, it was injected onto a hydroxyapatite column and eluted with an ammonium sulphate gradient. Again, a single peak was obtained containing pure T7 polymerase.

The protein was concentrated to 10 mg/ml and then diluted to 5 mg/ml with 100% glycerol (to obtain a mixture containing 50% glycerol). It was flash frozen in liquid nitrogen and stored at -80˚C until needed. About 140 mg of protein were obtained from the 2 litres of culture.

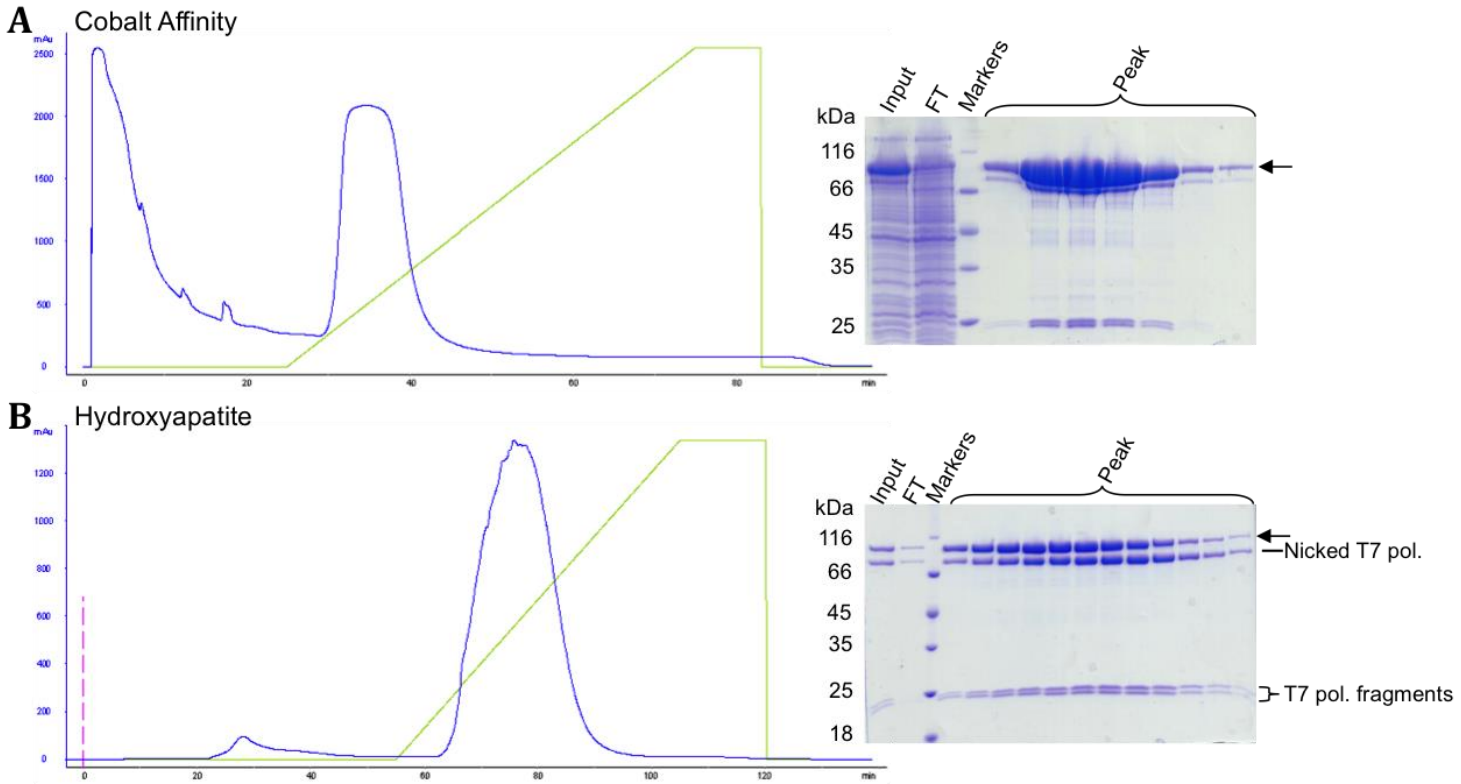


Figure 4.2: Purification of the T7 polymerase using a previously established protocol.

All chromatograms show absorbance at 280 nm in blue. The pink dotted line corresponds to the injection of the sample. The appropriate elution gradient is shown as a green line. The protein of interest is indicated on the gels with an arrow.

A) Affinity purification using cobalt resin. The protein bound to the column via its His tag and was eluted using an imidazole gradient (from 5 to 500 mM over 100 ml). A single peak that saturated the UV lamp was eluted. SDS-PAGE of the fractions showed that the peak contained the correctly sized protein and that most of the input bound to the column.

B) Hydroxyapatite purification of the T7 polymerase eluted from the cobalt column. The vast majority of the protein bound to the column and was eluted in a single peak over a 0 to 7.5% ammonium sulphate gradient. The peak contained pure full-length and nicked T7 polymerase as shown on the gel.

The activity of the purified T7 polymerase was immediately tested alongside the MEGAshortscript kit (Ambion). The results (**Figure 4.3**) showed that the transcription was very similar using the two protocols.

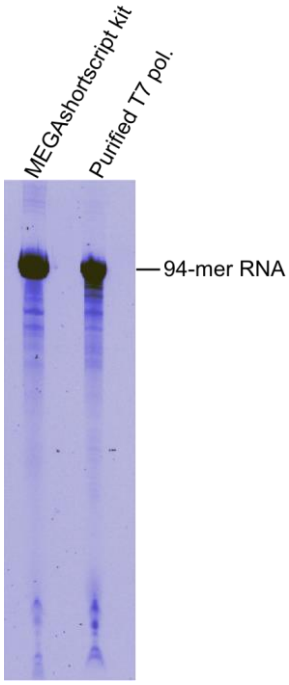


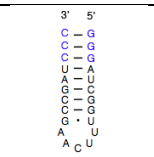
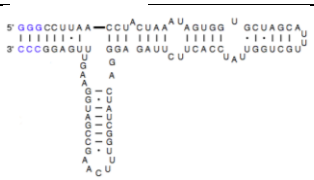
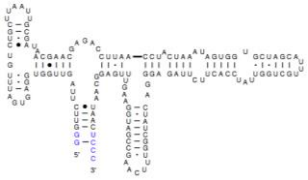
Figure 4.3: An *in vitro* transcription test using the commercial MEGAshortscript kit and the purified T7 polymerase.
Both methods show a high level of transcription in the same time. The commercial kit shows a slightly larger amount of the full-length RNA but the levels are comparable. Some degradation products are seen in both lanes and their amounts are also comparable.

4.2.2 Successfully transcribed helix 40 RNAs

The RNAs that were successfully transcribed are listed in **Table 4.1**. The DNA sequences were cloned into the pUC19 plasmid and linearized with restriction enzymes. This allowed the polymerase to reach the end of the template and fall off, terminating transcription. Alternative restriction sites were also present to allow the creation of a 3′ extension for primer extension (see section 4.3.2). The RNA was then transcribed using the MEGAshortscript kit (Ambion). **Figure 4.4** shows different RNAs on a denaturing gel.

Table 4.1: RNA sequences obtained from *in vitro* transcription and the structures predicted to form.

The helix 40 sequences from *S. cerevisiae* (ScH40) were successfully transcribed. Black letters in the sequences mark the regions that correspond to actual rRNA. Extra bases were added for more efficient cloning, transcription, and/or loop formation. These are shown in blue on the structure images and in the sequences. The structures shown are predicted to form *in vivo*.

Organism	RNA Name	Predicted Secondary Structure	RNA Sequence
S. cerevisiae	Short ScH40		GGGAUCGGUUUCAAGCCGAUCCC
S. cerevisiae	Intermediate ScH40		GGGCCUUAACCUACUAAAUAGUGGUGCUAGCAUUUGCUGGUUAUCCACUUCUUAGAGGGACUAUCGGUUUCAAGCCGAUGGAAGUUUGAGGCCC
S. cerevisiae	Long ScH40		GGGUUCUUAGUUGGUGGAGUGAUUUGUCUGCUUAAUUGCGAUAACGAACGAGACCUUAACCUACUAAAUAGUGGUGCUAGCAUUUGCUGGUUAUCCACUUCUUAGAGGGACUAUCGGUUUCAAGCCGAUGGAAGUUUGAGGCAAUAACUCCC

Additional *S. cerevisiae* and *C. thermophilum* short H40 RNAs were synthesized by IDT. These are shown in **Table 4.2**. They did not contain the extra bases (denoted in blue in **Table 4.1**) and were tested in some assays.

Table 4.2: Helix 40 RNA sequences from the two organisms synthesized by IDT.

They form the structure shown for the short H40 in Table 4.1 but without the extra blue bases. The sequences differ from each other by two nucleotides (marked in red). These lie in the single-stranded part and thus, do not affect the structure of the RNA.

RNA	Sequence
ScH40	AUCGGUUUCAAGCCGAU
CtH40	AUCGGCUCAAGCCGAU

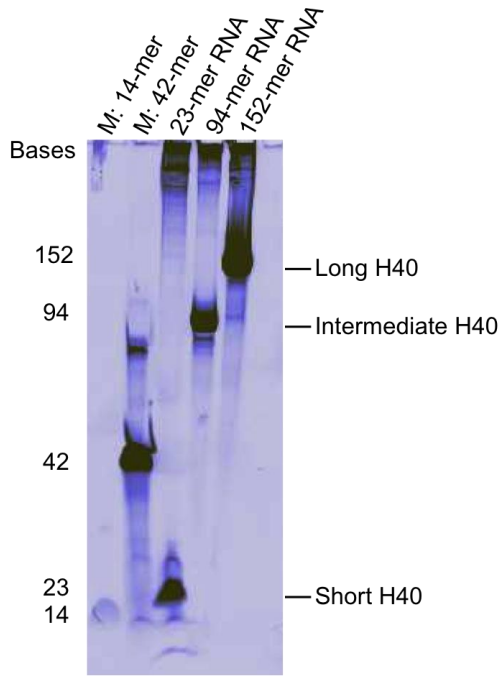


Figure 4.4: *In vitro* transcription of the different H40 RNAs cloned.
Two marker RNAs (marked with an M) were run on the gel as size standards (14 and 42 bases long). The transcribed RNAs ran at the appropriate sizes although some products were stuck in the wells. This could be due to secondary structures not being fully denatured or due to the plasmid not being completely digested and some long RNAs being made. Either way, the gel shows that the H40 RNAs can be made and purified.

4.2.3 Determination of helix 40 RNA secondary structure by chemical modification and primer extension

The structure of the RNA likely plays a role in its ability to bind to Nob1. Since the RNA is taken out of the ribosomal environment, its structure could be affected. Thus, the secondary structure of the short constructs was probed using chemical modification followed by primer extension.

RNAs were transcribed with an extended 3′ vector sequence to allow the binding of the primer. They were then subjected to treatment with 1M7, which modifies flexible nucleotides usually indicating non-base pairing. Reverse transcription was used to identify the modified (i.e. single-stranded) regions as the enzyme stopped transcribing more often at these bases. The different DNA fragments were then separated on a denaturing gel along with sequencing ladders. This allowed the mapping of the modified regions and the secondary structure of the RNAs to be confirmed.

Figure 4.5 shows the results of the primer extension. The two shorter RNAs were confirmed to have the expected structure. The 5′ end of the long RNA could not be clearly visualised on the gel. A second, internal primer would be needed to clarify this region. However, the parts that could be analysed confirmed the predicted structure of the important helix 40.

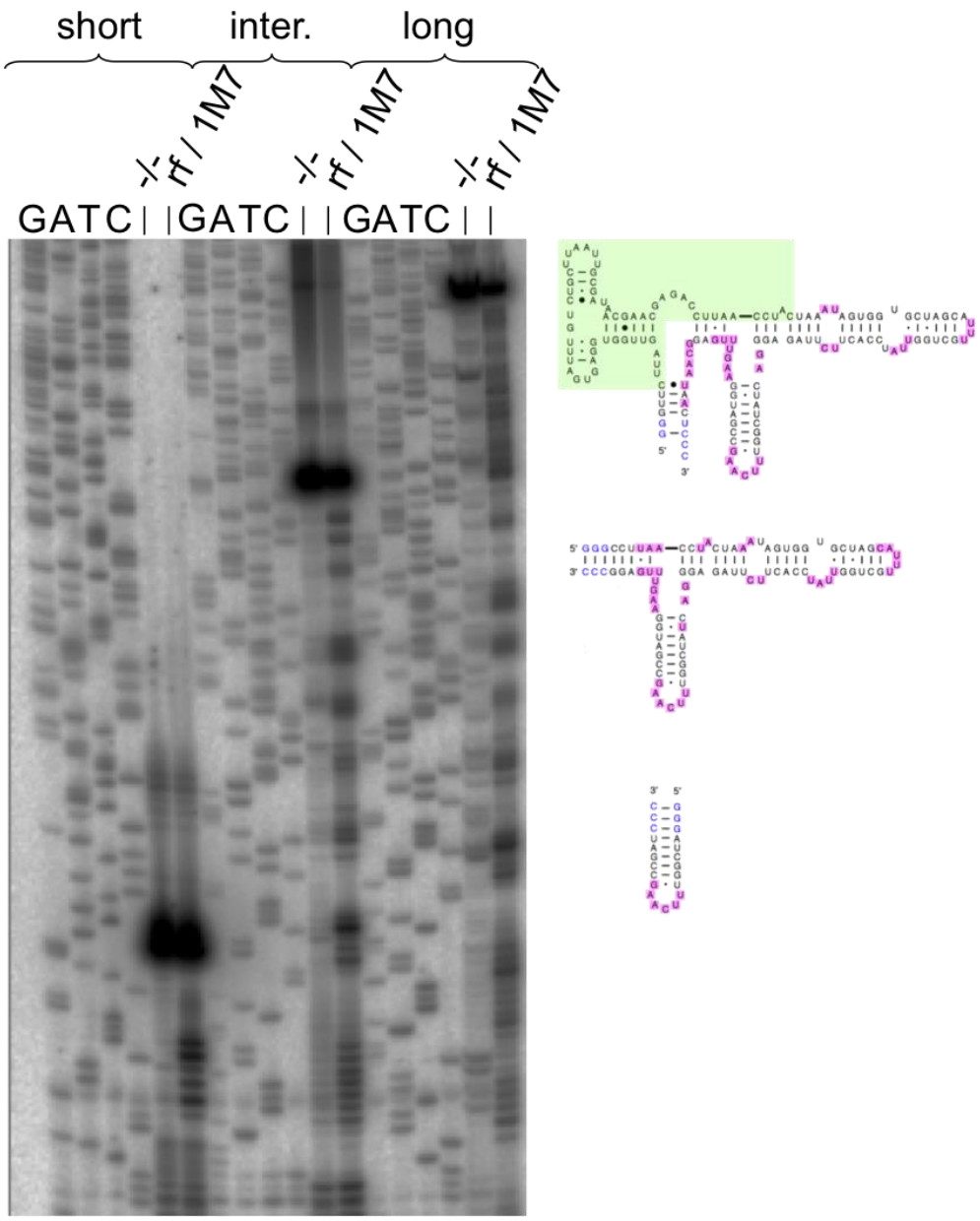


Figure 4.5: Primer extension results using the three different H40 RNAs.
Each of the transcribed fragments was subjected to 1M7 (1-methyl-7-nitroisatoic anhydride) modification. This reagent only reacts with 'open' regions (i.e. single-stranded regions). When a base shows this modification, it causes polymerases to stop prior to replicating it. Thus, when the fragments were then subjected to reverse transcription using a radioactively labelled (5′ ^{32}P label) primer, fragments of certain lengths were found to be more abundant (i.e. when the reverse transcriptase paused due to the modification). This allowed mapping of the single-stranded regions. Each refolded and modified fragment (rf / 1M7 lane) was separated on a 6% denaturing gel along with the corresponding sequencing ladder and a sample without refolding or 1M7 modification (-/- lane) to allow mapping of natural stops. A phosphoscreen was exposed to the gel and then scanned. The result can be seen here and are summarised on the picture to the right (predicted structures). Blue letters represent bases that were added to aid in hairpin formation but are not found in the rRNA sequence. Pink shows the bases that were shown to be single-stranded during primer extension. The green region on the long RNA could not be accurately mapped as the bases are too close together at the top of the gel. Overall, the structures appear to be correct, especially around helix 40 (the helix pointing down).

The assay was repeated with ScNob1 in the mixture to observe whether protein binding caused helix 40 RNA to be protected from modification. However, there was no change in the observed pattern.

4.2.4 Neither CtNob1 nor ScNob1 interact with Helix 40 RNA

The interaction between the short H40 RNA hairpin and CtNob1 or ScNob1 was tested in semi-analytical SEC and EMSAs. Upon complex formation, a larger species is expected to form. This would correspond to a shift to the left in SEC (i.e. an earlier elution volume). Also, the shifted peak would be expected to have a higher absorbance, as it would correspond to protein and RNA. In the EMSAs, the larger complex would migrate more slowly through a native gel and, thus, appear as a higher band.

In SEC, the transcribed short hairpin and the synthetic *C. Thermophilum* (CtH40) RNAs were used. **Figure 4.6** shows that the ScNob1 does not interact with ScH40 in these conditions. The same was seen using CtNob1 and CtH40 (**Figure 4.7**) or ScH40 (**Figure 4.8**).

Finally, the CtNob1 protein and the intermediate ScH40 RNA were tested. However, as shown in **Figure 4.9**, the elution profile for the RNA alone was very complex and could not be easily analysed. Nevertheless, no major changes were seen when the protein was added. Therefore, it is unlikely that there was an interaction. This was confirmed by analysing the fractions on a silver stained SDS-PAGE gel. The two species did not co-elute and remained separate.

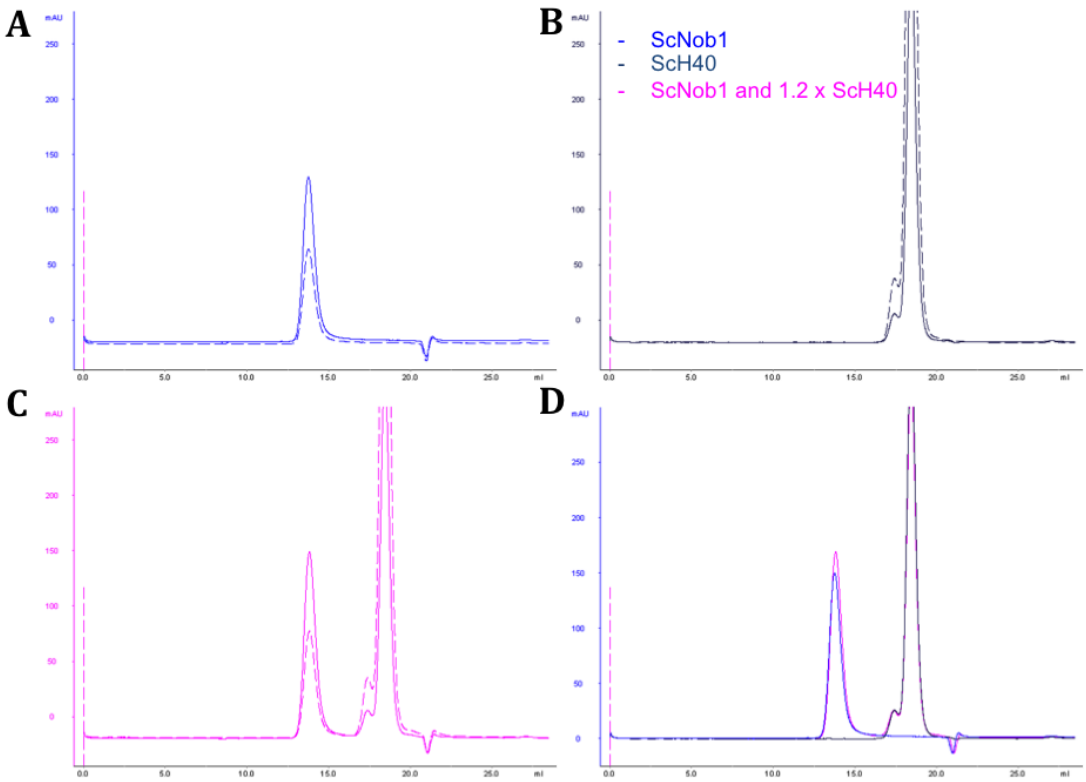


Figure 4.6: Semi-analytical SEC of ScNob1 and ScH40.
Absorbance at 280 nm is shown as a solid line whereas absorbance at 260 nm is shown as a dashed line. The key for the colours is shown on the right.

A) The SEC profile for ScNob1 only. A single peak is seen.

B) The chromatogram for ScH40 RNA alone. Two overlapping peaks are seen, likely due to the presence of some RNA dimers as well as the hairpin.

C) The SEC result when ScNob1 is mixed with ScH40 in a 1:1.2 ratio. The peaks remain separate and look like an additive effect.

D) An overlay of the three previous results showing absorbance only at 280 nm. The tops of the peaks are in the same positions on the elution profile suggesting that there is no interaction. A slight increase in absorbance can be seen. This is likely due to pipetting or injection variation.

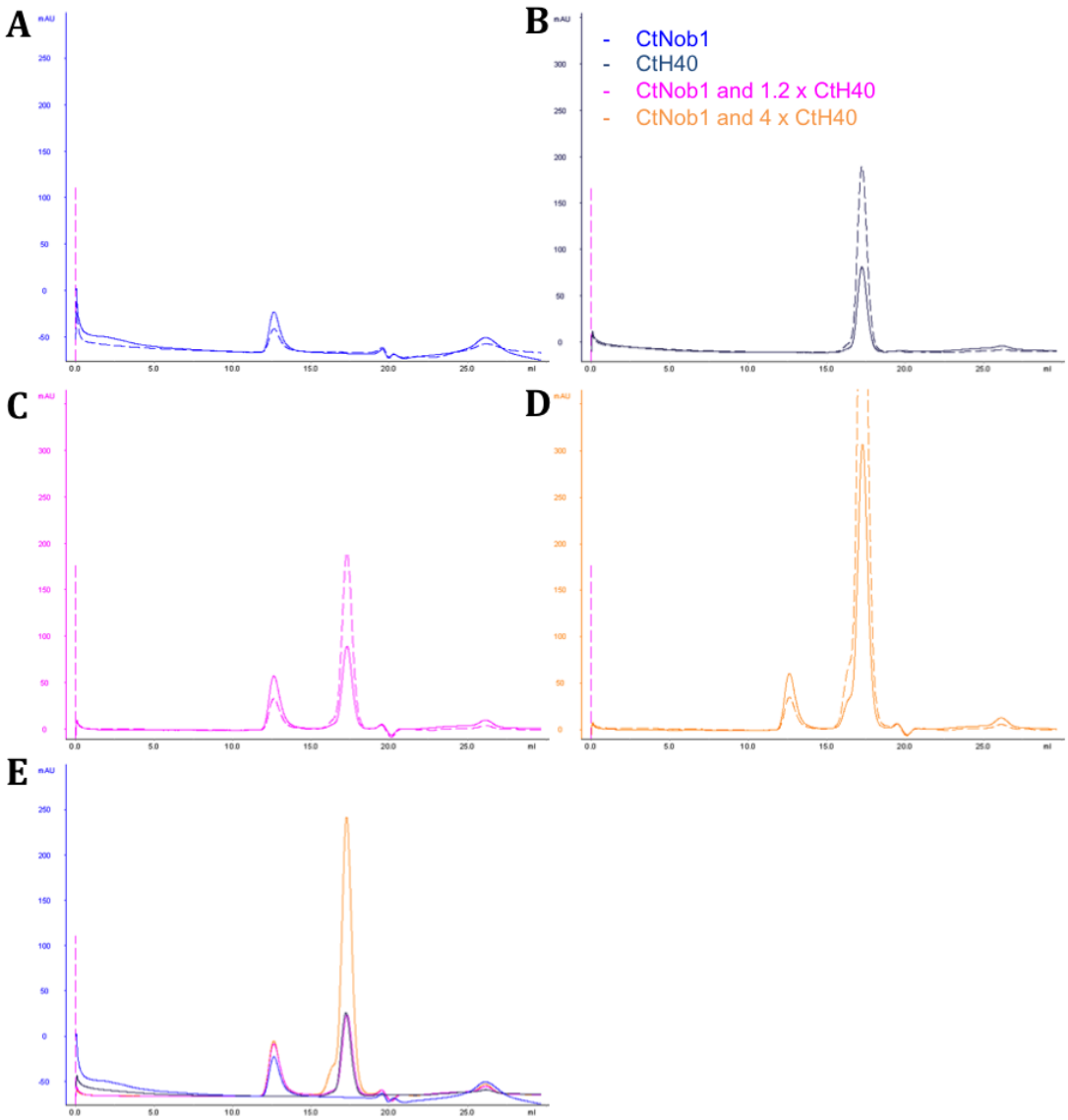


Figure 4.7: Semi-analytical SEC of CtNob1 and CtH40.
Absorbance at 280 nm is shown as a solid line whereas absorbance at 260 nm is shown as a dashed line. The key for the colours is shown on the right.

A) The SEC profile for CtNob1 only. A small single peak is observed.

B) The chromatogram for CtH40 RNA alone. A single peak is seen suggesting that no RNA dimers are formed with this RNA.

C) The SEC result when CtNob1 is mixed with CtH40 in a 1:1.2 ratio. The peaks remain separate like in the previous chromatograms.

D) Results for CtNob1 and CtH40 mixed in a 1:4 ratio to verify that there are no interactions. The peaks still remain separate although a 'bump' is seen in the RNA peak suggesting that at higher concentrations, RNA dimers are formed.

E) An overlay of the four previous results showing absorbance only at 280 nm. The tops of the peaks are, again, in the same positions on the elution profile suggesting that there is no interaction. Some changes in the height of the peaks are seen due to experimental variations.

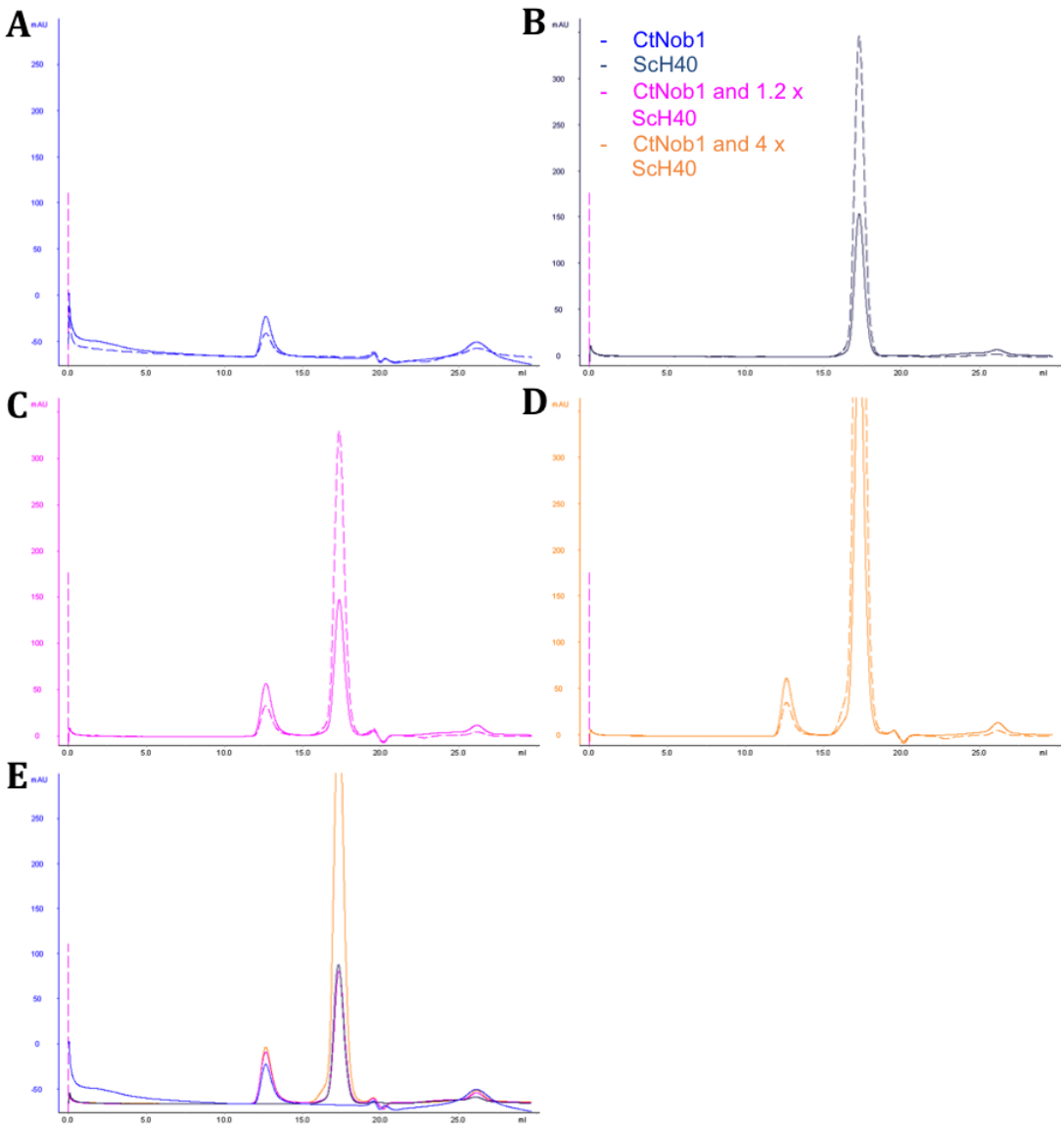


Figure 4.8: Semi-analytical SEC of CtNob1 and ScH40.
Absorbance at 280 nm is shown as a solid line whereas absorbance at 260 nm is shown as a dashed line. The key for the colours is shown on the right.

A) The SEC profile for CtNob1 only. A small single peak is observed.

B) The chromatogram for ScH40 RNA alone. A single peak is seen suggesting that no RNA dimers are formed this time.

C) The SEC result when CtNob1 is mixed with ScH40 in a 1:1.2 ratio. The peaks remain separate and look like the individual chromatograms. The slight changes in sequence appear to have no effect.

D) Results for CtNob1 and ScH40 mixed in a 1:4 ratio. The peaks still remain separate. Again, a bump is seen on the RNA peak suggesting dimer formation at higher concentrations.

E) An overlay of the four previous results showing absorbance only at 280 nm. The tops of the peaks are, as with the other experiments, in the same positions on the elution profile suggesting that there is no interaction. The sequence differences between ScH40 and CtH40 do not give any changes in the results.

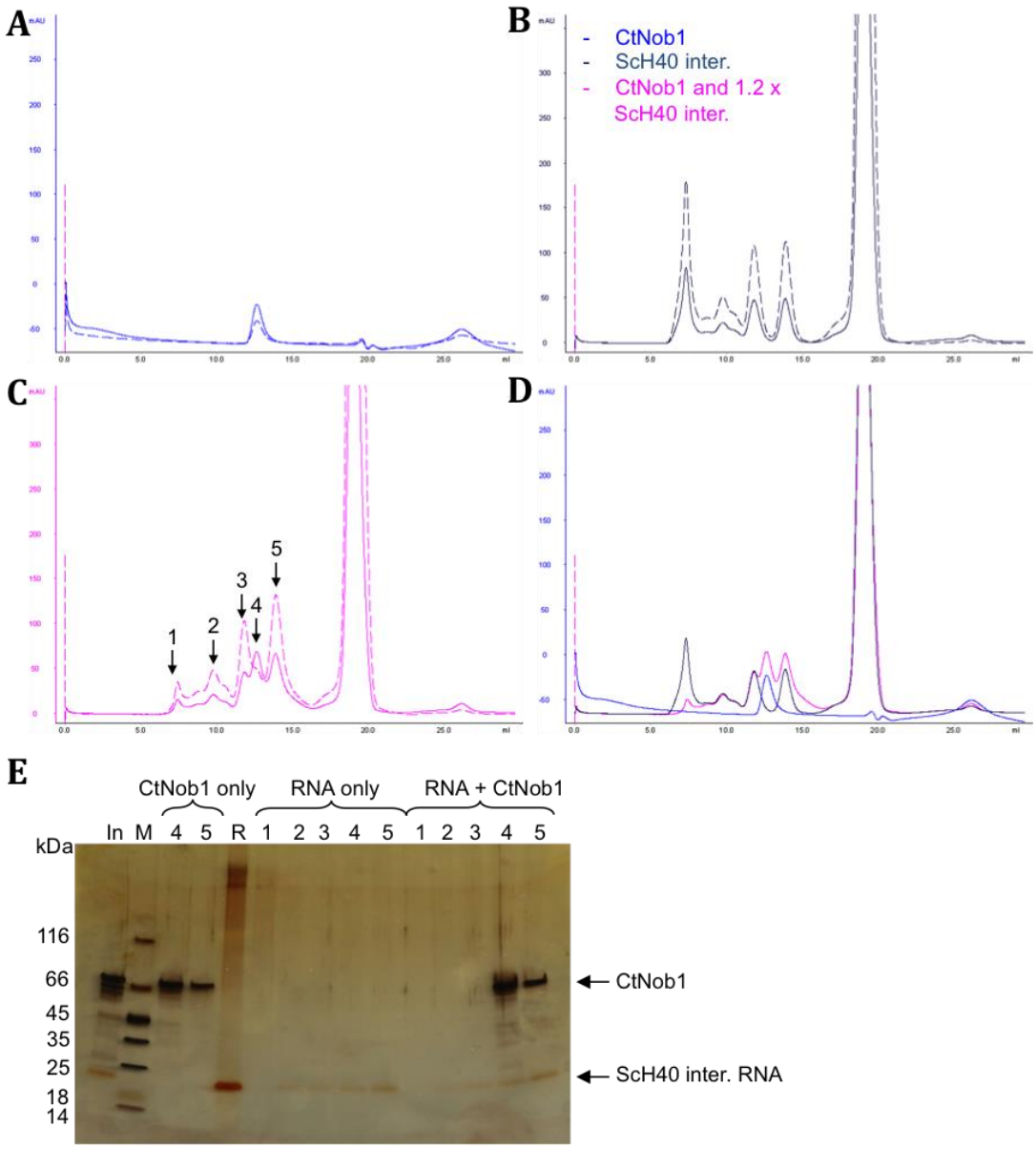


Figure 4.9: Semi-analytical SEC of CtNob1 and ScH40 intermediate RNA.
Absorbance at 280 nm is shown as a solid line whereas absorbance at 260 nm is shown as a dashed line. The key for the colours is shown on the right.

A) The SEC profile for CtNob1 only. A small single peak is observed.

B) The chromatogram for ScH40 intermediate RNA alone. Multiple peaks are seen suggesting that many different conformations of the RNA exist in the same sample.

C) The SEC result when CtNob1 is mixed with ScH40 intermediate RNA in a 1:1.2 ratio. The peaks remain in the same positions although the heights of some of them do change.

D) An overlay of the three previous results showing absorbance only at 280 nm. The tops of the peaks are in the same positions on the elution profile suggesting that there is no interaction. Some changes in the height of the peaks are seen. These could be caused by an interaction. However, this is unlikely as the positions (i.e. sizes/shapes of the molecules) stay the same.

E) Silver stained gel of the fractions covering the peaks labelled in image **C)**. The lanes include the input (In), markers (M), numbered peaks (1-5), and just RNA (R). The position and distribution of the RNA and protein do not change suggesting that there is no interaction.

For the EMSAs, the synthetic H40 sequences from IDT without any extra bases were used. The lack of any strong protein-RNA interactions was also observed in these experiments (**Figure 4.10**), confirming the observations obtained with *in vitro* transcribed sequences (which contained extra bases to stabilise the hairpin) used in SEC. For the ScNob1, the EMSA was repeated after the protocol was optimised (optimisation shown later). Therefore, the band for the complex appears in a different place on the gel (i.e. not in the well).

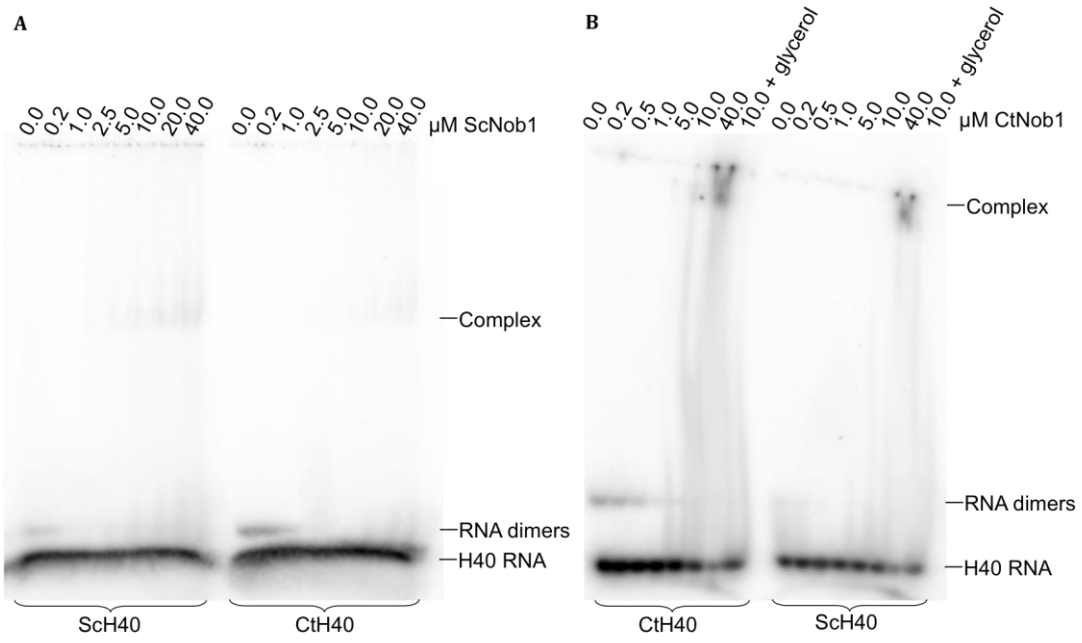


Figure 4.10: EMSA results for ScNob1 and CtNob1 binding to ScH40 and CtH40.

A) ScNob1 binding to ScH40 and CtH40. The amount of protein added to 0.2 µM RNA was varied as seen on top of the image. The amount of free ScH40 RNA is not seen to decrease significantly. However, some complex formation is seen at 10-40 µM protein. The CtH40 experiment shows slightly less complex forming suggesting that the sequence might be important. However, the interaction is very weak in both cases. Thus, it is potentially completely non-specific.

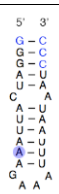
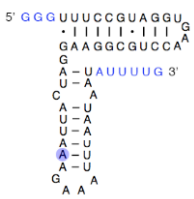
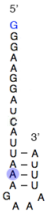
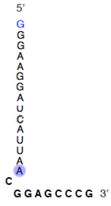
B) CtNob1 binding to CtH40 and ScH40. In the last reaction, 30% glycerol was added to see if it would affect binding (as glycerol was present in the analytical gel filtration experiment). No difference was seen in amount of free CtH40 or ScH40 present when glycerol is added. Again, some complex is seen to form. However, this is extremely inefficient and could be due to non-specific interactions. This is supported by the fact that both sequences behave in the same manner.

4.3 Nob1 Does Interact with D-site Derived Sequences

4.3.1 *In vitro* transcribed D site RNAs

In order to test whether the Nob1 protein can interact with the D site RNA sequence, several constructs were created for *in vitro* transcription. Some of the RNAs were directly transcribed from DNA sequences, while others were synthesized by IDT. **Table 4.3** lists some of the different RNAs transcribed using the purified T7 polymerase. Other RNAs used are shown later in the relevant sections.

Table 4.3: Some of the different D site RNAs transcribed.
Black letters in the sequences mark the regions that correspond to actual rRNA. Extra bases were added for more efficient cloning, transcription, and/or loop formation. These are shown in blue on the structure images and in the sequences. The red adenine is where the D site cleavage occurs. This site is also shown on the structure images in a blue circle. The first two sequences were cloned into the pUC19 plasmid for transcription while the lower two were transcribed from annealed DNA oligonucleotides. The structures shown are predicted to form *in vivo*, although there is still some debate about whether the D site is double- or single- stranded.

Organism	RNA Name	Predicted Secondary Structure	RNA Sequence
S. cerevisiae	D site 0 (DS0)		GGGAUCAUUAAAGAAAUUUAAUAAUCCC
S. cerevisiae	Long D site		GGGUUUCCGUAGGUGAACCUGCGGAAGGAUCAUUAAAGAAAUUUAAAUUUUG
S. cerevisiae	D site 1 (DS1)		GGGAAGGAUCAUUAAAGAAAUUUA
H. sapiens	D site 1 (DS1)		GGGAAGGAUCAUUAACGGAGCCCG

4.3.2 ScNob1 does not interact with DS0 but does with DS1

Initially, the yeast DS0 and DS1 RNAs from **Table 4.3** were tested with ScNob1 to see if an interaction could be observed. Semi-analytical SEC was performed and showed that the ScNob1 did not bind to DS0 in these conditions (**Figure 4.11**). However, as seen in **Figure 4.12**, when DS1 (which is predicted to form a partial hairpin) was used, a complex was formed, suggesting that a single-stranded substrate is required for binding.

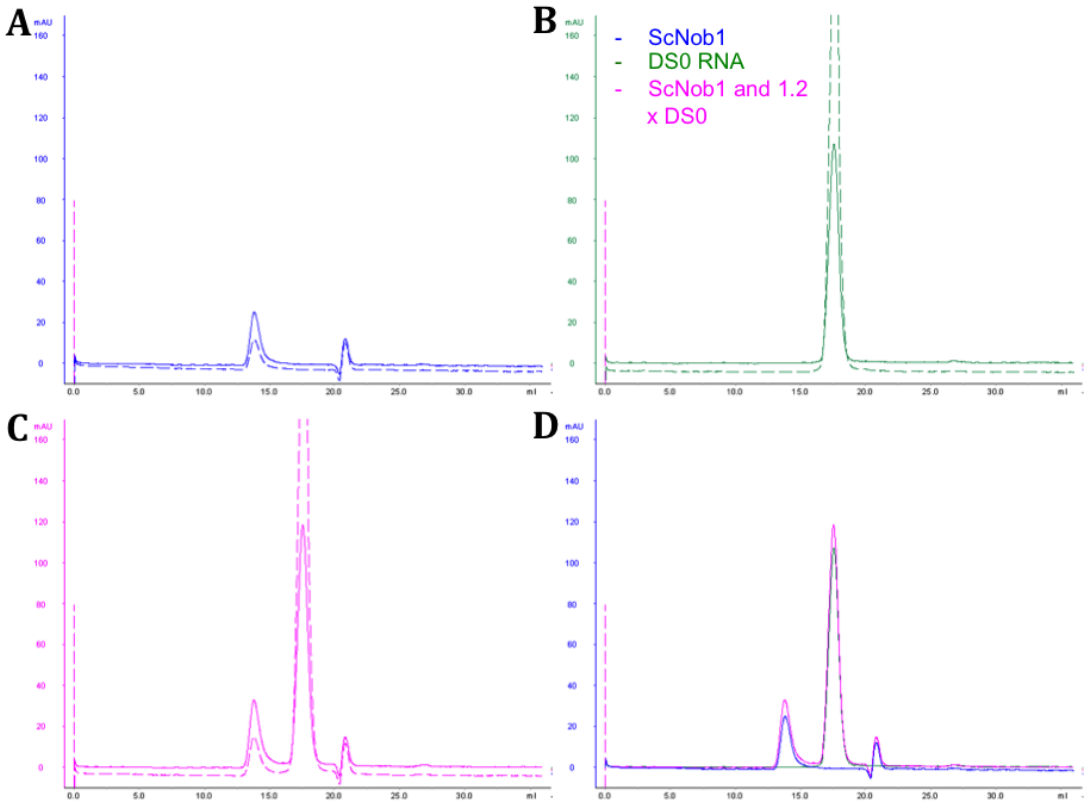


Figure 4.11: Semi-analytical SEC of ScNob1 and DS0 RNA.
Absorbance at 280 nm is shown as a solid line whereas absorbance at 260 nm is shown as a dashed line. The key for the colours is shown on the right.

A) The SEC profile for ScNob1 only. A single peak is seen.

B) The chromatogram for DS0 RNA alone. A single large peak is seen showing that a single RNA species is present.

C) The SEC result when ScNob1 is mixed with DS0 in a 1:1.2 ratio. The peaks remain separate and look like the individual profiles, suggesting an additive effect.

D) An overlay of the three previous results showing absorbance only at 280 nm. The peaks are in the same positions on the elution profile suggesting that there is no interaction. A slight increase in absorbance can be seen. This is likely due to pipetting or injection variation.

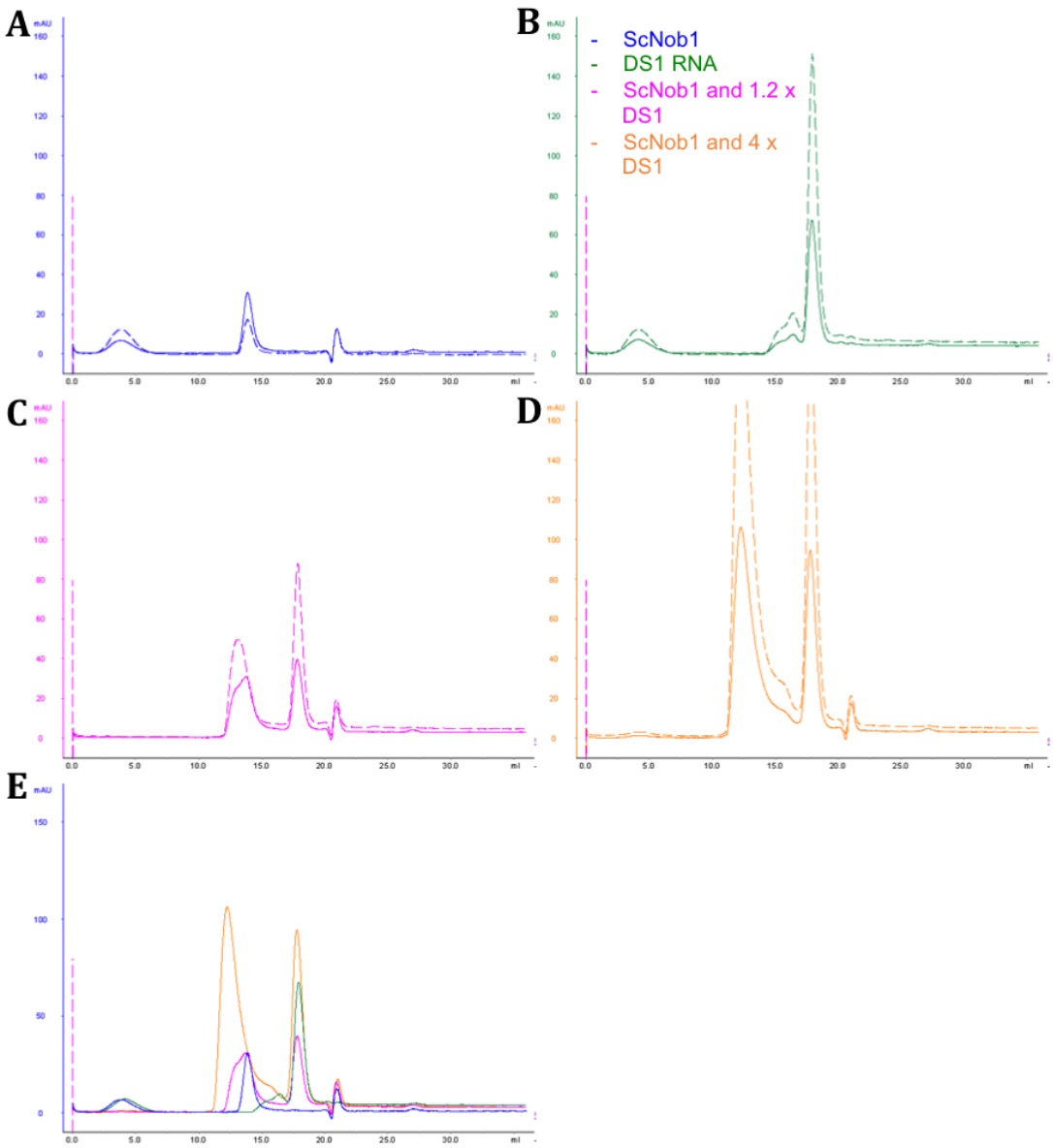


Figure 4.12: Semi-analytical SEC of ScNob1 and DS1.
Absorbance at 280 nm is shown as a solid line whereas absorbance at 260 nm is shown as a dashed line. The key for the colours is shown on the right.

A) The SEC profile for ScNob1 only. A small single peak is observed.

B) The chromatogram for DS1 RNA alone. Multiple peaks are seen showing that different conformations exist.

C) The SEC result when ScNob1 is mixed with DS1 in a 1:1.2 ratio. There is a slight shift in the position of the peak and an increase in absorbance at 260 nm suggesting complex formation.

D) Results for ScNob1 and DS1 mixed in a 1:4 ratio. A much larger shift to an earlier elution volume is seen, again, supporting that a complex is formed.

E) An overlay of the four previous results showing absorbance only at 280 nm. The increase in size and absorbance when the DS1 RNA is added is clearly seen.

These results were then confirmed with an EMSA. However, when a standard EMSA protocol based on a TBE buffer system (described in **Table 4.4**) was used (**Figure 4.13**), the complex would not enter the gel, possibly due to the pH of the TBE buffer (8.3) being below the pI of the protein (8.5). Thus, the protein was not negatively charged and was retained in the well.

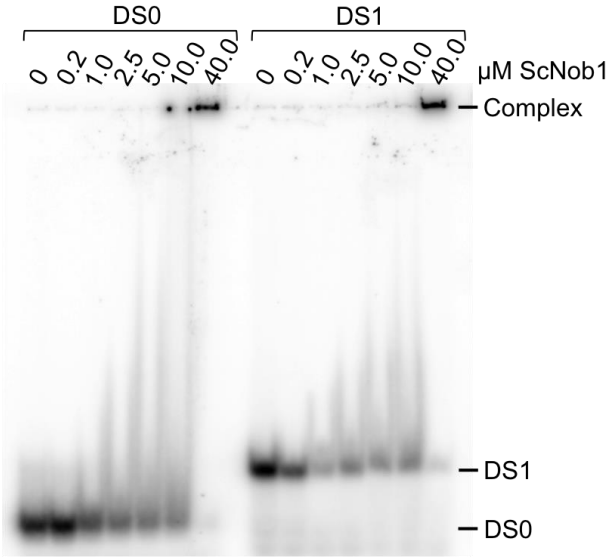


Figure 4.13: Initial standard EMSA using ScNob1 and DS0 or DS1 RNA.
0.2 μM RNA was used with increasing amounts of protein. The free RNA can be seen at the bottom of the gel while the complex gets stuck in the well. Both RNAs are seen to form a complex although more of the free DS1 RNA is absent at lower protein concentrations.

To optimise the protocol, several buffer systems were tested. Most showed no improvement. The best two buffers (at pH 9.4 and 10.2 shown in **Table 4.4**), described by McLellan (1982), were used on non-radioactive non-denaturing PAGE gels with ScNob1 and the DS0 RNA (as it showed retention of protein in the wells using the standard EMSA protocol). As shown in **Figure 4.14**, both allowed the protein to enter the gel. However, some protein was still present in the wells on the pH 9.4 gel. Thus, the ammonia/CAPS buffer at pH 10.2 was chosen. **Figure 4.15** shows the results of the EMSA using this system.

Table 4.4: The different buffers used for EMSAs.
The standard names of the buffers and final concentrations of the components are shown. CAPS is also known as 3-(Cyclohexylamino)-1-propanesulfonic acid. The first entry is the standard buffer used for EMSAs. Other buffers were also tested but showed no improvement.

Buffer name	pH	Basic component	Acidic component	Additional component
0.5 x TBE	8.3	45 mM Tris base	45 mM Borate	1 mM EDTA
1.0 x TC	9.4	60 mM Tris base	40 mM CAPS	-
1.0 x AC	10.2	37 mM Ammonia	20 mM CAPS	-

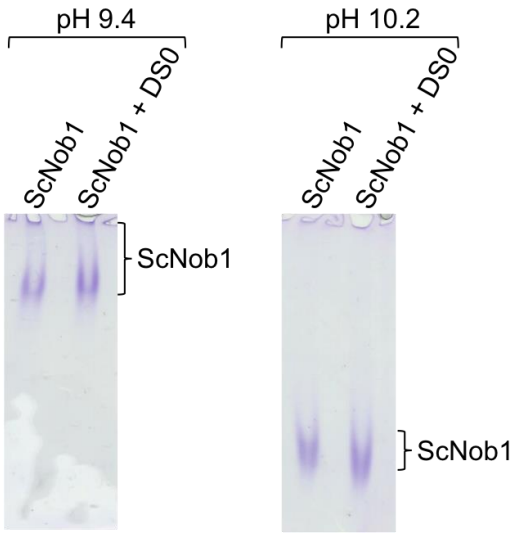


Figure 4.14: PAGE analysis of ScNob1 without and with DS0 RNA using different EMSA buffer systems.
A non-denaturing acrylamide gel containing the tested buffer was run for each system. It was then stained with Coomassie dye to detect the ScNob1. At pH 9.4 the protein is able to enter the gel although some is still staining at the well. At pH 10.2 all of the protein is seen to enter the gel in a relatively sharp band.

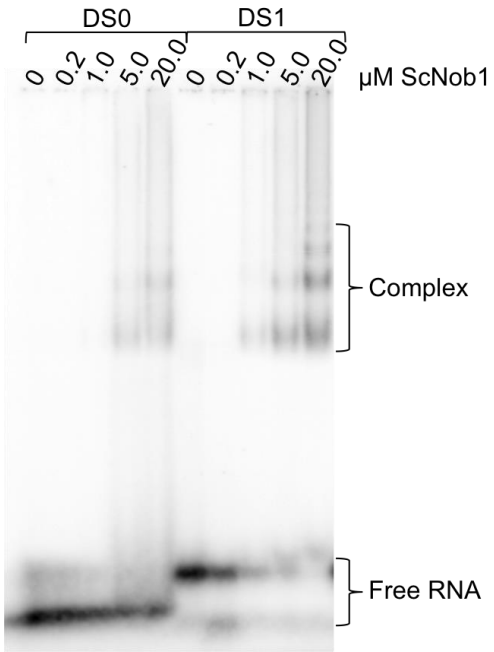


Figure 4.15: Optimised EMSA using ScNob1 and DS0 or DS1.
Again, 0.2 µM RNA was used with increasing protein concentrations. As before, both RNAs are seen to form a complex. Free DS0 RNA is still seen using a 100 times excess of protein. DS1 RNA starts to disappear at a 1:1 ratio and is completely absent at 5.0 µM ScNob1. With higher protein concentrations, higher order structures are seen (i.e. multiple bands for the complex are present).

Some complex formation with DS0 at high ScNob1 concentrations is seen. However, there is also free RNA present at these concentrations. Thus, it seems that the hairpin RNA is not able to bind to the ScNob1 protein efficiently. All of the free DS1 RNA is retarded at 10-50 µM ScNob1 suggesting much stronger binding than that observed with DS0.

4.3.3 ScNob1 binding site overlaps with the predicted ScDim2 site

In order to narrow down the ScNob1 binding site on the D site RNA, extensive EMSA experiments using different sequences (listed in **Table 4.5**) were performed. The optimised protocol was used with each of the RNAs. The EMSAs were later repeated with a new protein preparation and the results remained the same. **Figure 4.16** shows all the EMSAs performed and the results are summarised in **Table 4.6**.

Table 4.5: RNAs synthesized by IDT for EMSA experiments.
The first nucleotide removed by D-site cleavage is shown in red. The predicted Dim2 binding site is shown in green and extra bases are shown in blue. DS0 and DS1 are the same as the previously described *in vitro* transcribed sequences.

Name	Sequence (5′ - 3′)
DS0	GGGAUCAUUAAAGAAAUUUAAUAAUCCC
DS1	GGGAAGGAUCAUUAAAGAAAUUUA
DS2	GAAGGAUCAUUAAAGAAAUUUAAUAAU
DS3	GAAGGAUCAUUAAAGAAAUUUA
DS4	GAACCUGCGGAAGGAUCAUUAAAGAAAUUUA
DS5	GGGUUUCCGUAGGUGAACCUGCGGAAGGAUCAUUAAAGAAAUUUAAUAAU
DS6	GGGAAGGAUCAUUAAAGAAA
DS7	GGGAAGGAUCAUUAAAG
DS8	GGAUCAUUAAAGAAAUUUAAUAAU
DS9	GGAUCAUUAAAGAAAUUUA
DS10	AUUAAAGAAAUUUAAUAAU

The results suggest that binding to the D site is most strongly affected when the predicted Dim2 binding site (see Chapters 1 and 5 for details) is removed. Thus, it is possible that the binding sites for the two proteins overlap and that only one protein can bind at any time.

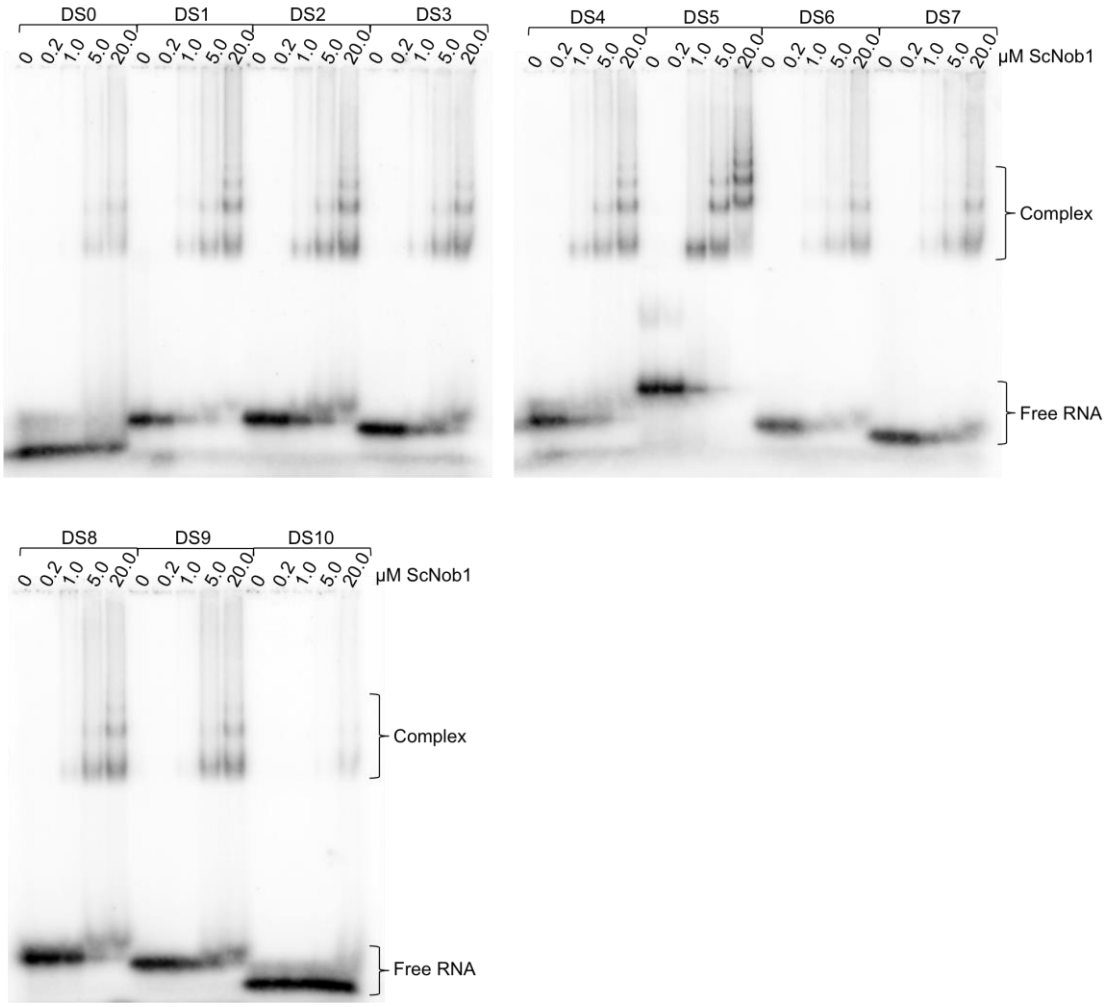
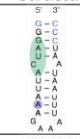
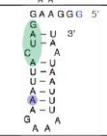
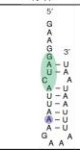
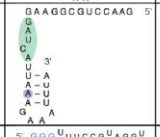
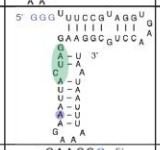
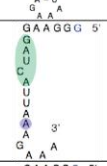
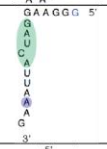
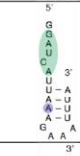


Figure 4.16: EMSA results for DS0-DS10 RNAs with ScNob1.
The relative amounts of complex formed using each of the RNAs (at 0.2 μM) can be seen. The RNAs that do interact with ScNob1 show increased levels of higher order structures as the protein concentration is increased. All of the RNAs are seen to bind apart from DS10, which shows no decrease in free RNA levels. A small amount of complex is still seen at a 100 times excess of protein but that is likely due to non-specific interactions.

Table 4.6: Summary of EMSA results using DS0-DS10 with ScNob1.

The colouring is the same as in the previous table. The Dim2 binding site is marked with a green text in the sequences and a green oval in the structure images. The binding to ScNob1 is scored based on the amount of complex observed in the EMSAs. DS0 was assigned a score of 1 while DS1 was assigned a score of 3. All other RNAs were scored based how the complex formation compared to these two RNAs.

Name	Sequence	Predicted Secondary Structure	ScNob1 binding
DS0	GGGAUCAUUAAAGAAAUUUAAUAAUCCC		1
DS1	GGGAAGGAUCAUUAAAGAAAUUUA		3
DS2	GAAGGAUCAUUAAAGAAAUUUAAUAAU		2-3
DS3	GAAGGAUCAUUAAAGAAAUUUA		2-3
DS4	GAACCUGCGGAAGGAUCAUUAAAGAAAUUUA		3
DS5	GGGUUUCCGUAGGUGAACCUGCGGAAGGAUCAUUAAAGAAAUUUAAUAAU		3
DS6	GGGAAGGAUCAUUAAAGAAA		2
DS7	GGGAAGGAUCAUUAAAG		2
DS8	GGAUCAUUAAAGAAAUUUAAUAAU		2
DS9	GGAUCAUUAAAGAAAUUUA		1
DS10	AUUAAAGAAAUUUAAUAAU		0

4.3.4 The AGGA sequence is the binding determinant

From the EMSA experiments, it seemed that ScNob1 interacted with the D site in the same location as Dim2. To further verify the location of the interaction, ITC experiments were performed using RNAs with point mutations. During a binding event, heat is released or absorbed (depending on whether the reaction is exo- or endothermic respectively). In ITC, these heat changes are measured during consecutive injections of one reactant into the other. The data is integrated and thermodynamic values and dissociation constants are calculated from the resulting binding curve.

Although DS0 and DS1 were tested in the ITC (**Figure 4.17**), the mutant RNAs were based on DS6. This RNA still binds to the protein but lacks any predicted secondary structures that might cause heat changes in the ITC. The different mutations tested can be seen in **Table 4.7**.

Table 4.7: RNAs synthesized by IDT for ITC experiments.
The first nucleotide removed by D-site cleavage is shown in red. The predicted Dim2 binding site is shown in green and extra bases are shown in blue. Mutated nucleotides are shown in orange.

Name	Sequence (5′ - 3′)
DS0	GGGAUCAUUAAAGAAAUUUAAUAAUCCC
DS1	GGGAAGGAUCAUUAAAGAAAUUUA
DS6 (Sc)	GGGAAGGAUCAUUAAAGAAA
DS6.1	GGGAGGGAUCAUUAAAGAAA
DS6.2	GGGAAAGAUCAUUAAAGAAA
DS6.3	GGGAAGAAUCAUUAAAGAAA
DS6.4	GGGAAAAAUCAUUAAAGAAA
DS6.5	GGGAAGGGUCAUUAAAGAAA
DS6.5.2	GGGAAGGUUCAUUAAAGAAA
DS6.6	GGGAAGGAACAUUAAAGAAA
DS6.7	GGGAAGGAUUAUUAAAGAAA
DS6.8	GGGAAGGAUCGUUAAAGAAA
DS6.9	GGGAAGGAUCAUUGAAGAAA
DS6.10	GGGAAGGAUCAUUAGAGAAA
DS6.11	GGGAAGGAUCAUUAAAGAGG
DS11	GGGAAGGAUCAUUA
DS12	GGGAUCAUUAAAGAAA
DScram1	GAAAGGAGAGGAUUACAAUA
DScram2	GAAAAAAGCGGUAGUAAUAG
DS6 (Hs)	GGGAAGGAUCAUUAACGGAG

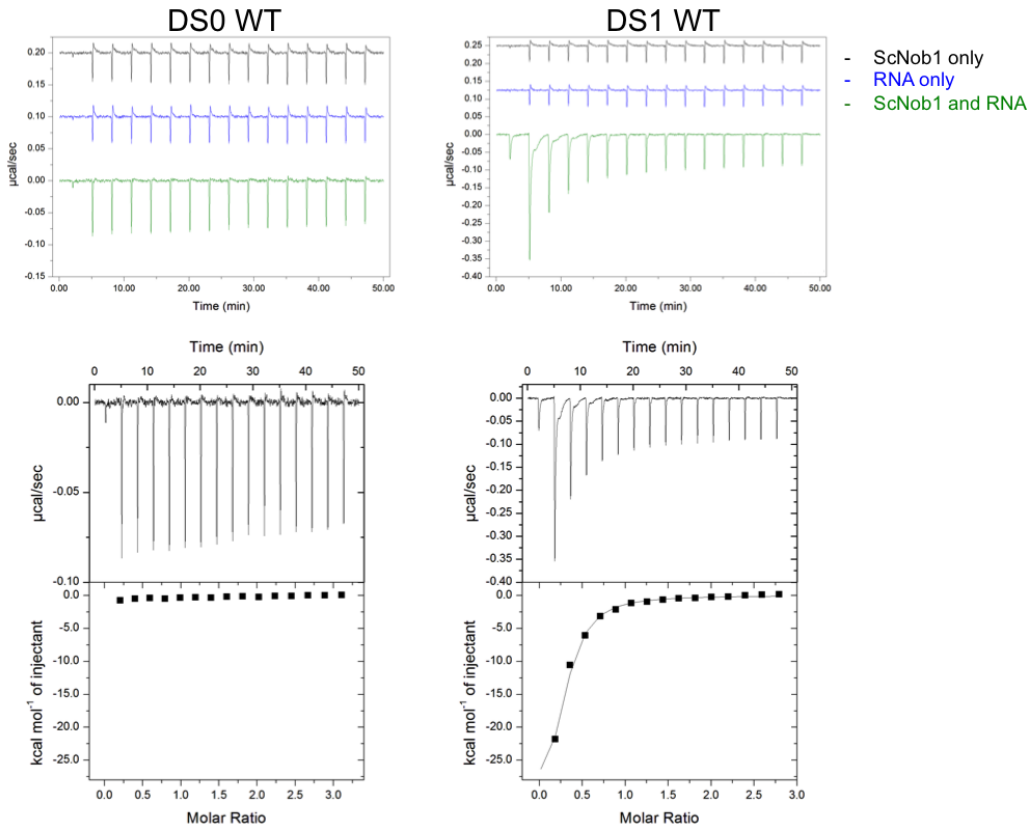


Figure 4.17: ITC results for DS0 and DS1 RNAs binding to ScNob1.
The raw data for the protein control (black), RNA control (blue), and the experiment (green) are shown in the top panel. The protein and RNA controls were subtracted from the experimental data to give the lower graph. The peaks were integrated and plotted as a binding curve for each of the RNAs (shown as the bottom panel). As seen above, the DS0 RNA does not bind to ScNob1 and, thus, no significant heat changes are seen. The integrated binding curve looks like a straight line supporting that there is no binding. Conversely, the DS1 RNA is able to bind and shows a clear binding curve. The thermodynamic values obtained from the curve can be seen in **Table 4.9**. The molar ratio (0.25) shows that four RNA molecules were able to bind one ScNob1 molecule.

The WT DS6 was tested along with 11 mutants (DS6.1-11). Then, an RNA that mimics the cleaved product (DS11) was used. DS12 has the same starting sequence as DS0 but lacks the 3′ end, which creates the hairpin in DS0. A human version of the DS6 was also tested as it only differs in the 3′ end after the cleavage site. Finally, two scrambled RNAs were also included as controls (DScram1 and 2).

Figures 4.18-20 show the ITC results for the RNAs. The dissociation constant (K_d) values obtained from the experiments are shown in **Table 4.8**. It appears that the AGGA sequence that overlaps the predicted Dim2 binding site is the only part that affects ScNob1 binding. Point mutations in the AGGA sequence did

not completely abolish binding. However, the protein binding sites were not saturated suggesting a weaker interaction. Also, the binding event with the mutants appears more complex as there is a kink in the curves. A double mutant was tested (DS6.4) and that showed an additive effect: a small heat change was seen, but the binding was much weaker than for the individual point mutations.

DS11 binds better than DS6 suggesting that the RNA is not efficiently released after the cleavage and might require additional factors. The first scrambled RNA did bind to the protein. However, this RNA was found to contain an AGGA sequence. When this was removed in DScram2, the RNA was no longer able to bind.

Additional thermodynamic values were also obtained and are listed in **Table 4.8**. However, these values varied using different RNAs. This can also be observed in the binding titrations as the magnitudes of the heat changes observed changed significantly with different sequences due to unknown factors. For example, the changes in heat and enthalpy (ΔH) observed depended on the heat of folding of RNA/protein as well as the binding event (Feig, 2007). These different contributions made it impossible to distinguish differences in enthalpy caused by binding alone. Thus, these values were not used for more specific analysis.

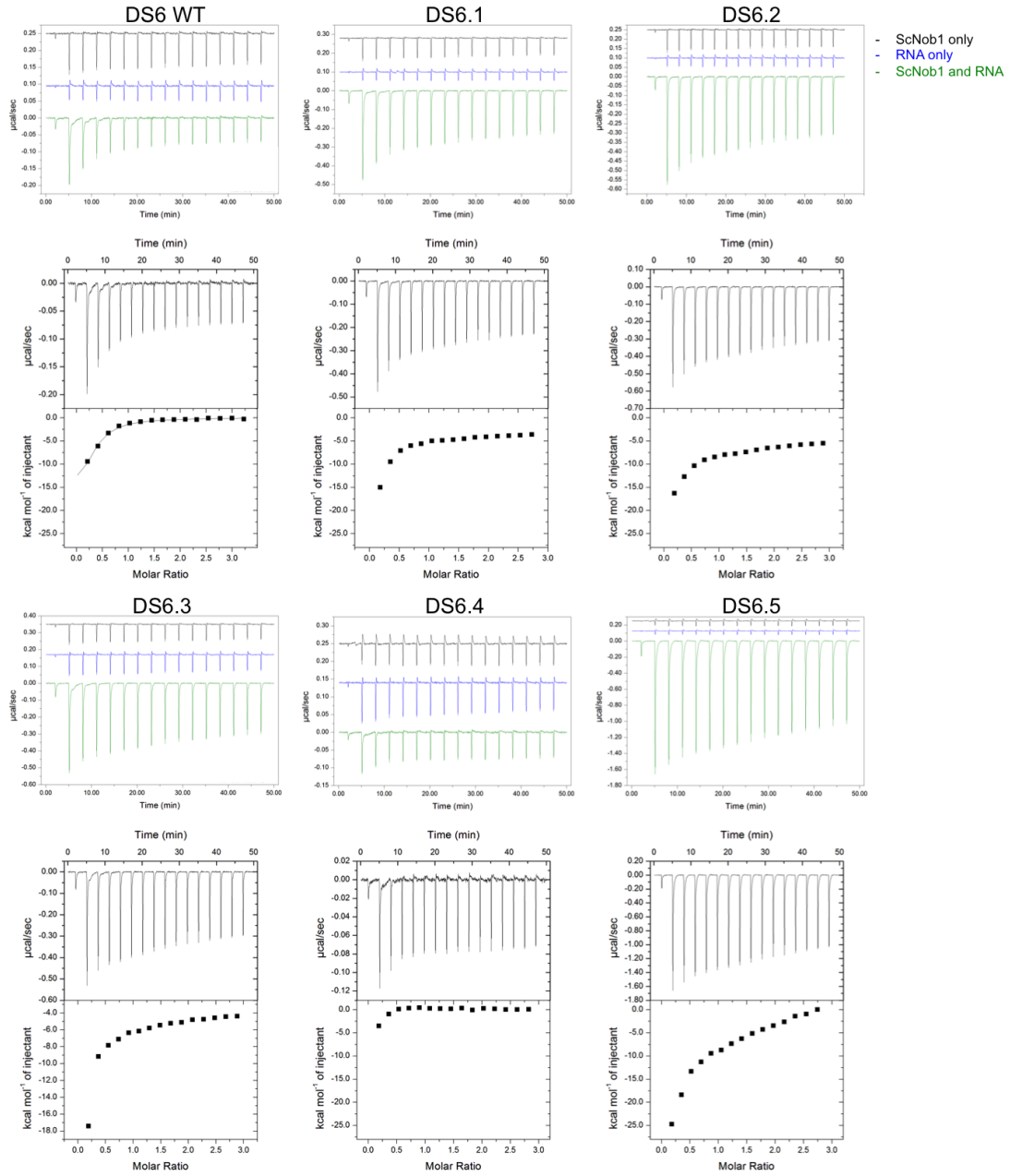


Figure 4.18: ITC results for the DS6 RNA and the 6.1-5 mutants.
All binding curves are shown on the same scale as for DS1 in the previous figure. The WT DS6 RNA shows a full binding curve that comes to saturation at the end. The molar ratio corresponds to 0.25 as with DS1. Again, all thermodynamic values can be seen in **Table 4.8**. The values for all of the mutants shown in this image could not be determined. Mutants 6.1, 2, 3, and 5 showed incomplete curves due to a lower affinity for the substrate. Also, they showed a kink in the curve suggesting a more complex reaction. Mutant 6.4 showed a small dip in the binding curve. However, this was too small to allow a curve to be fitted implying that no significant interaction is seen.

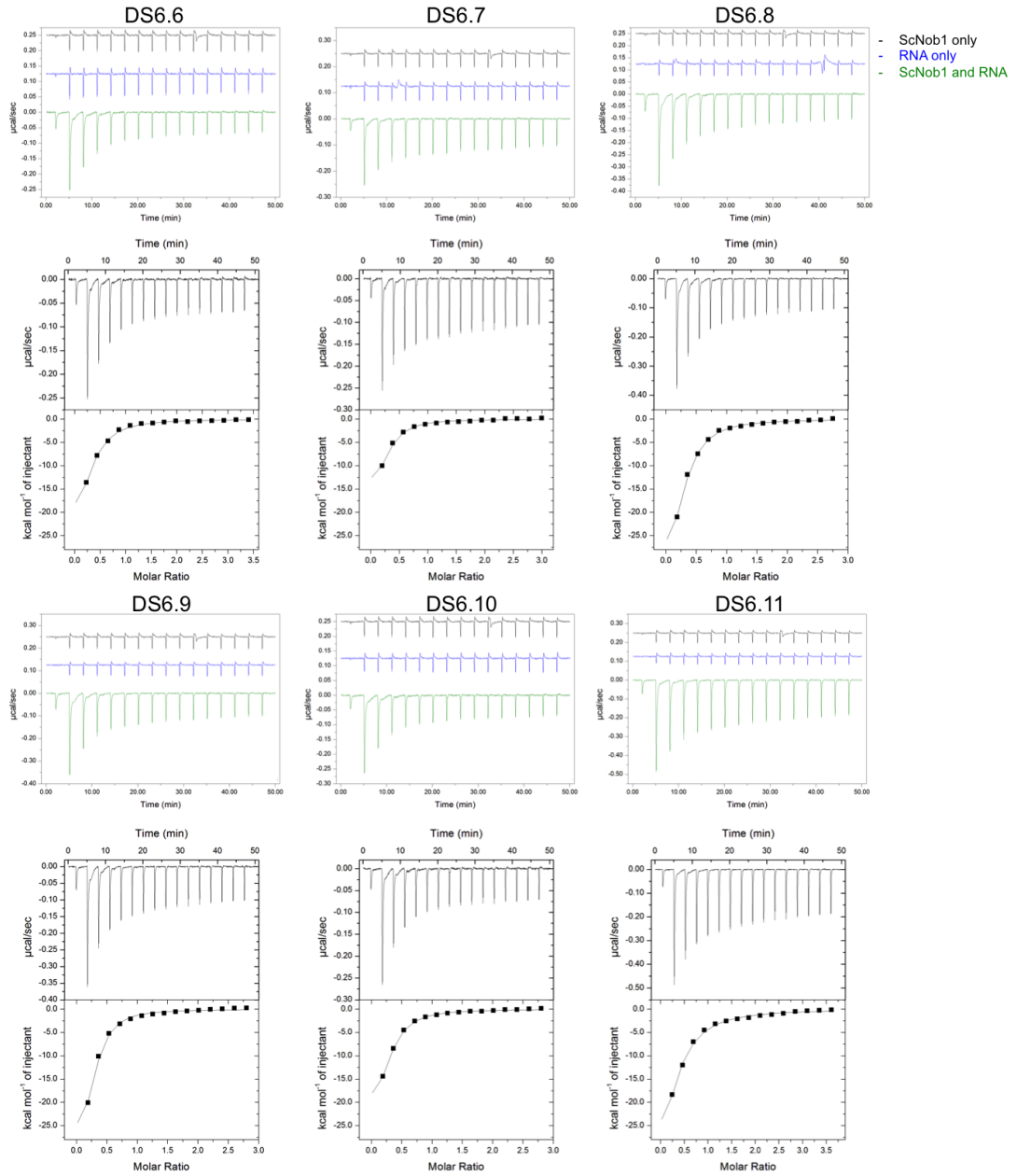


Figure 4.19: ITC results for mutants DS6.6-11.
All of the mutants shown here were able to bind to the ScNob1 protein showing similar binding curves to the WT DS6. The magnitude of the heat changes does vary even though the dissociation constants are similar. As with the others, the thermodynamic values can be found in **Table 4.8**.

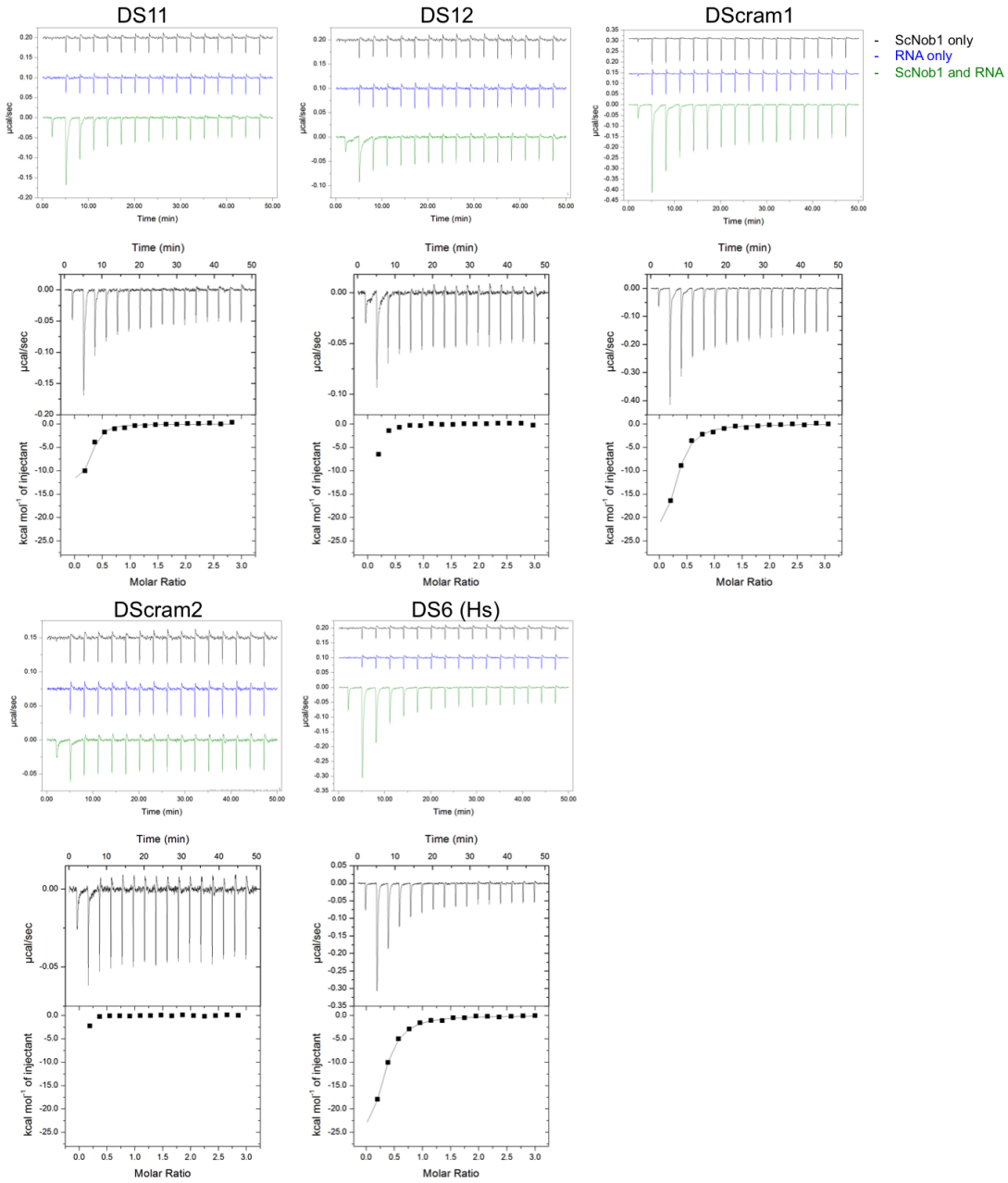


Figure 4.20: ITC results for DS11, DS12, DScram1-2, and the human DS6 sequence with ScNob1.

DS11 mimics the product of the D site cleavage and is able to bind to Nob1 as shown here. DS12 has the same sequence as DS0 but does not have any predicted secondary structures. It shows small heat changes. However, these changes are too small to fit a curve. DScram1 is a scrambled sequence that contains AGGA within it. It shows binding which is comparable to DS6. Conversely, DScram2 lacks the AGGA and does not appear to interact with the ScNob1 protein to any significant extent. Finally, the human DS6 sequence, which differs in the 3′ sequence, is able to bind to ScNob1 as well as the yeast one.

Table 4.8: Summary of the thermodynamic values obtained from the ITC experiments for each of the RNAs tested.
The RNAs without any values could not be fitted with a simple one-site binding model and/or did not come to completion.

Name	K_d (µM)	ΔH (cal/mol)	ΔS (cal/mol.K)	TΔS (cal/mol)	ΔG (cal/mol)	ΔG (kcal/mol)
DS0	No fit	-	-	-	-	-
DS1	1.0	-36580	-104.0	-28912	-7668	-7.7
DS6	1.9	-13240	-21.4	-5949	-7291	-7.3
DS6.1	No fit	-	-	-	-	-
DS6.2	No fit	-	-	-	-	-
DS6.3	No fit	-	-	-	-	-
DS6.4	No fit	-	-	-	-	-
DS6.5	No fit	-	-	-	-	-
DS6.5.2	1.1	-23930	-58.7	-16319	-7611	-7.6
DS6.6	1.9	-33540	-94.3	-26215	-7325	-7.3
DS6.7	1.3	-19500	-43.2	-12010	-7490	-7.5
DS6.8	1.7	-42430	-126.0	-35028	-7402	-7.4
DS6.9	1.1	-34350	-96.1	-26716	-7634	-7.6
DS6.10	1.5	-28100	-74.2	-20628	-7472	-7.5
DS6.11	3.1	-60420	-192.0	-53376	-7044	-7.0
DS11	0.6	-12120	-14.8	-4114	-8006	-8.0
DS12	No fit	-	-	-	-	-
DScram1	1.3	-30210	-81.1	-22546	-7664	-7.7
DScram2	No fit	-	-	-	-	-
DS6 (Hs)	1.3	-35000	-98.7	-27439	-7561	-7.6

The RNAs used in the ITC had no predicted secondary structures. However, a native gel was run to see if any structures did form. **Figure 4.21** shows that some of the RNAs did indeed have more complex conformations. Thus, it is possible that some of the variation in the ability of ScNob1 to bind to the RNA and in enthalpy stemmed from the secondary structures. Breaking base interactions within hairpins and double strands consumes and releases a lot of energy causing large heat changes in the cell during the ITC run. These can mask changes caused by the protein-RNA interaction. It is noteworthy that the higher the concentration of the RNA, the more secondary structures are present. The concentrations in ITC were much lower than on the gel. Thus, it is likely that fewer structures existed in the ITC.

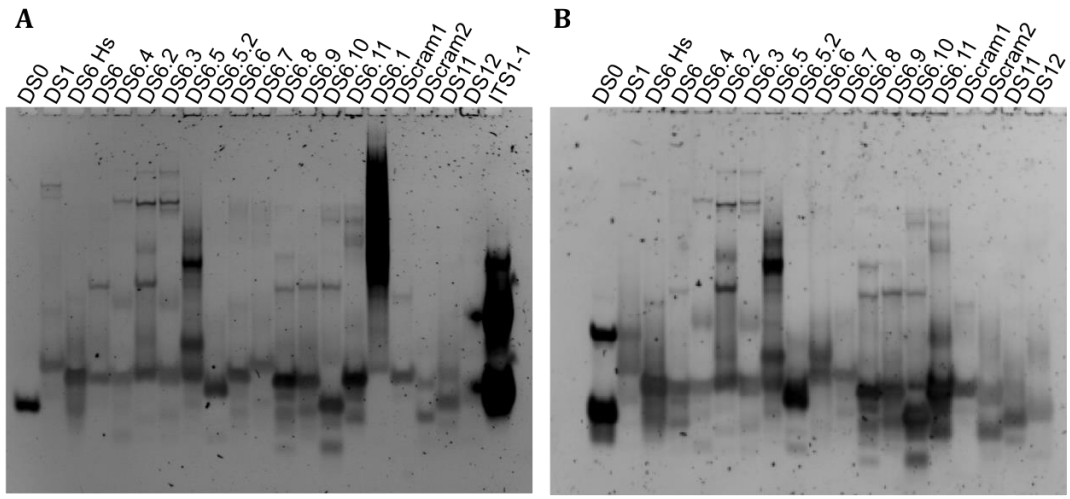


Figure 4.21: Native gels of the RNAs used in ITC.

A) A native gel stained with SYBR Safe showing all of the RNAs used for ITC. Most of the RNAs contain some secondary structures even though none were predicted. All of the RNAs showing changes in binding (DS6.1-5) show some amount of complex formation. The WT DS6 also shows a small amount of a species that runs higher on the gel.

B) A native gel without DS6.1 and ITS1 as they were overloaded in the previous gel preventing analysis of some of the other RNAs. Again, most of the RNAs show a mix of secondary structures. The amount of DS0 was increased to see if there is a concentration effect on the secondary structures. As shown here, the higher the concentration, the more secondary structures are present.

In order to test the effect of different conformations, mutant DS6.5.2 was created. This RNA has a mutation in the same site as DS6.5 but the adenine is changed to a uracil instead of a guanine. This mutation, as shown in **Figure 4.22**, rescues the ScNob1 binding. On the native gel, the RNA does not appear to have any secondary structures unlike DS6.5. Thus, it is possible that the RNA conformation plays a big role in ScNob1 binding. However, it is also possible that a smaller base is tolerated in this position.

Finally, all of the RNAs for which the binding curve could be determined showed a molar ratio of 0.25. This means that four RNA molecules were able to bind to one molecule of protein. This result was obtained with multiple protein batches. Repetition of the DS6-ScNob1 titration yielded the same results every time (average K_d of 1.8 ± 0.4 µM from four repeats).

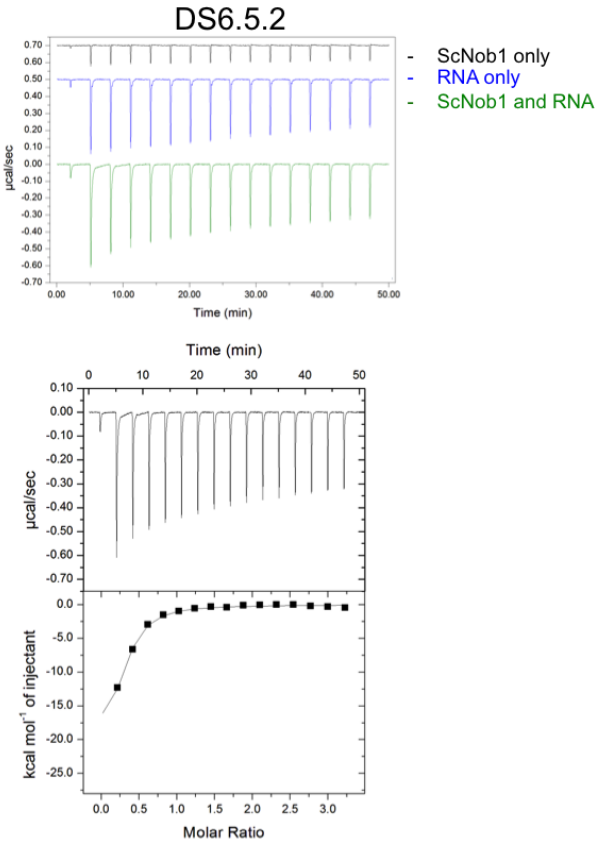


Figure 4.22: ITC results for DS6.5.2 binding to ScNob1.
This mutant shows the same binding curve as WT DS6. Thus, it appears that this mutation is able to rescue ScNob1 binding.

4.3.5 The stoichiometry of the interaction remains unclear

From previous experiments, it is clear that ScNob1 interacts with the D site RNA. However, it is unclear what the stoichiometry is. The semi-analytical gel filtration and EMSAs show an increase in size and ITC suggests a four RNAs to one protein ratio. The particles were then visualised with EM. However, as seen in **Figure 4.23**, the addition of RNA causes the formation of very large complexes.

A sample of protein with the DS1 RNA was sent to Dr. Leech in York to run on SEC-MALS. However, the sample eluted in the void suggesting a size of several MDa (data not shown). Again, this supports the formation of large complexes.

Finally, the same result could be observed in semi-analytical SEC of ScNob1 with higher concentrations of human DS1. The complex was seen to elute in many fractions including the void volume (**Figure 4.24**).

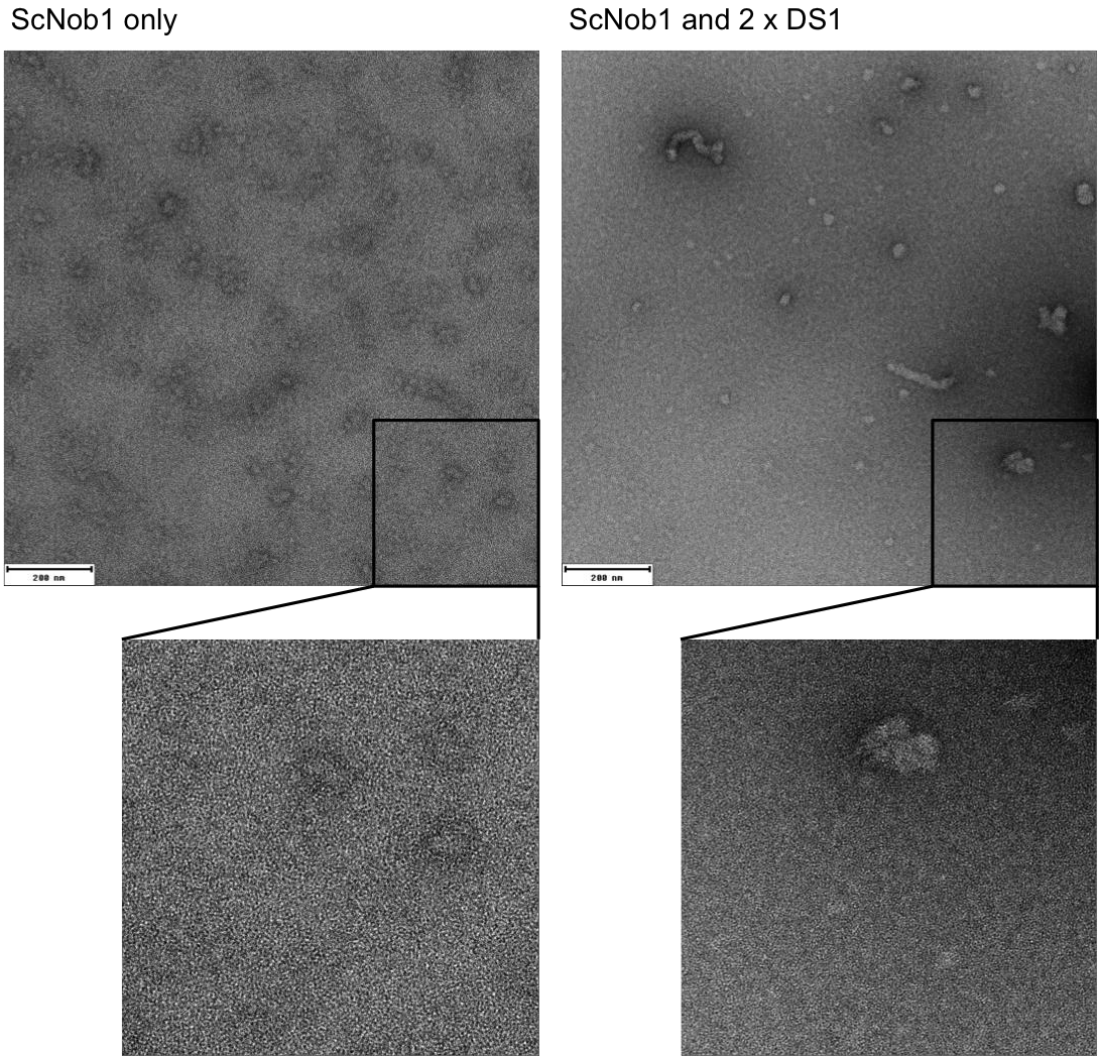


Figure 4.23: Example EM images of ScNob1 alone and with a 2 times excess of DS1 RNA.
The ScNob1 by itself shows well-distributed small particles in the negative stain. The addition of DS1 RNA causes the formation of larger, more distinct species. This suggests that multiple proteins and RNAs interact to make complex structures.

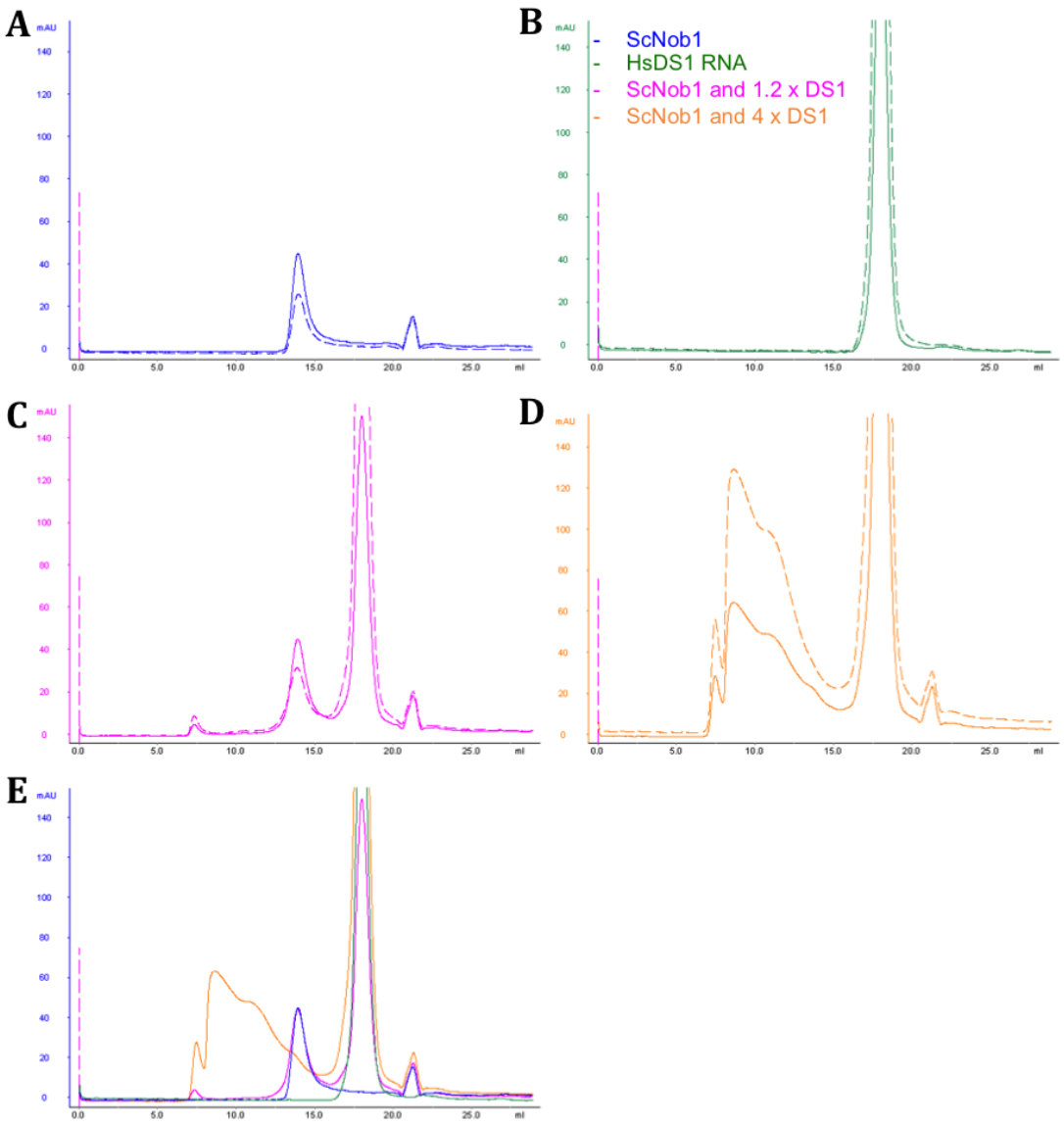


Figure 4.24: Semi-analytical SEC of ScNob1 and human DS1 RNA.
Absorbance at 280 nm is shown as a solid line whereas absorbance at 260 nm is shown as a dashed line. The key for the colours is shown on the right.

A) The SEC profile for ScNob1 only. A small single peak is observed.

B) The chromatogram for human DS1 RNA alone.

C) The SEC result when ScNob1 is mixed with DS1 in a 1:1.2 ratio. The main peaks remain in the same positions. However, some of the protein and RNA make a complex and elute in the void.

D) Results for ScNob1 and DS1 mixed in a 1:4 ratio. A much larger shift to an earlier elution volume is seen, again, supporting that a complex is formed. The complex shows a variety of sizes and elution volumes, including the void volume.

E) An overlay of the four previous results showing absorbance only at 280 nm. The increase in size when the DS1 RNA is added is clearly seen. Again, the size increase appears to be very significant.

4.4 Nob1 binds to an ITS1 loop sequence

The Tollervey lab has recently shown, using cross-linking *in vivo* in growing cells, that Nob1 interacts with a fragment of the internal transcribed spacer (ITS1) (unpublished data). The same site was identified *in vitro* by the Karbstein lab using chemical modification in the presence and absence of the protein (Lamanna & Karbstein, 2009). However, only a single base showed the lack of modification when Nob1 was present suggesting that this effect could be due to Nob1 proximity and not binding. Also, both studies utilised long, folded RNAs, which made data analysis more complex.

To test this theory using a short RNA, an ITS1 hairpin fragment (GGGCCUGCGCUUAAGUGCGCGGUCU, bases +96 to +120) was tested in ITC and semi-analytical SEC. The results in **Figures 4.25-26** show that the interaction can be seen *in vitro*. In ITC, the binding is a complex process and a simple one-site model could not be fitted to the binding curve. In SEC, all of the ScNob1 protein is saturated using a 1.2 times excess of RNA (unlike with DS1) suggesting that this interaction is stronger than the one with the D site and/or that only one binding site is present (unlike for the D site RNAs).

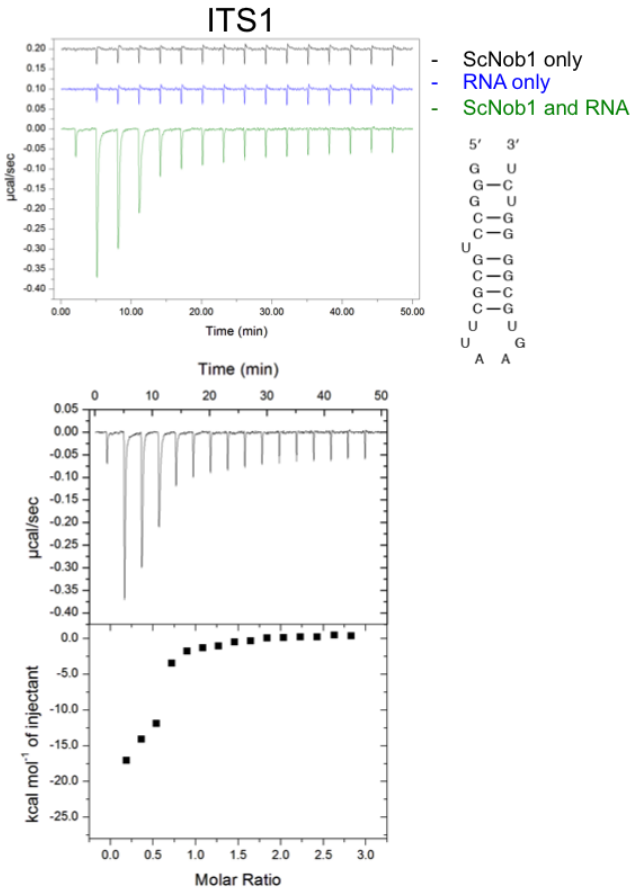


Figure 4.25: ITC results for the ITS1 fragment with ScNob1.
The predicted secondary structure of the RNA fragment is shown under the legend. A clear binding even can be seen. However, the curve does not show a simple one-site binding interaction and a model could not be fitted.

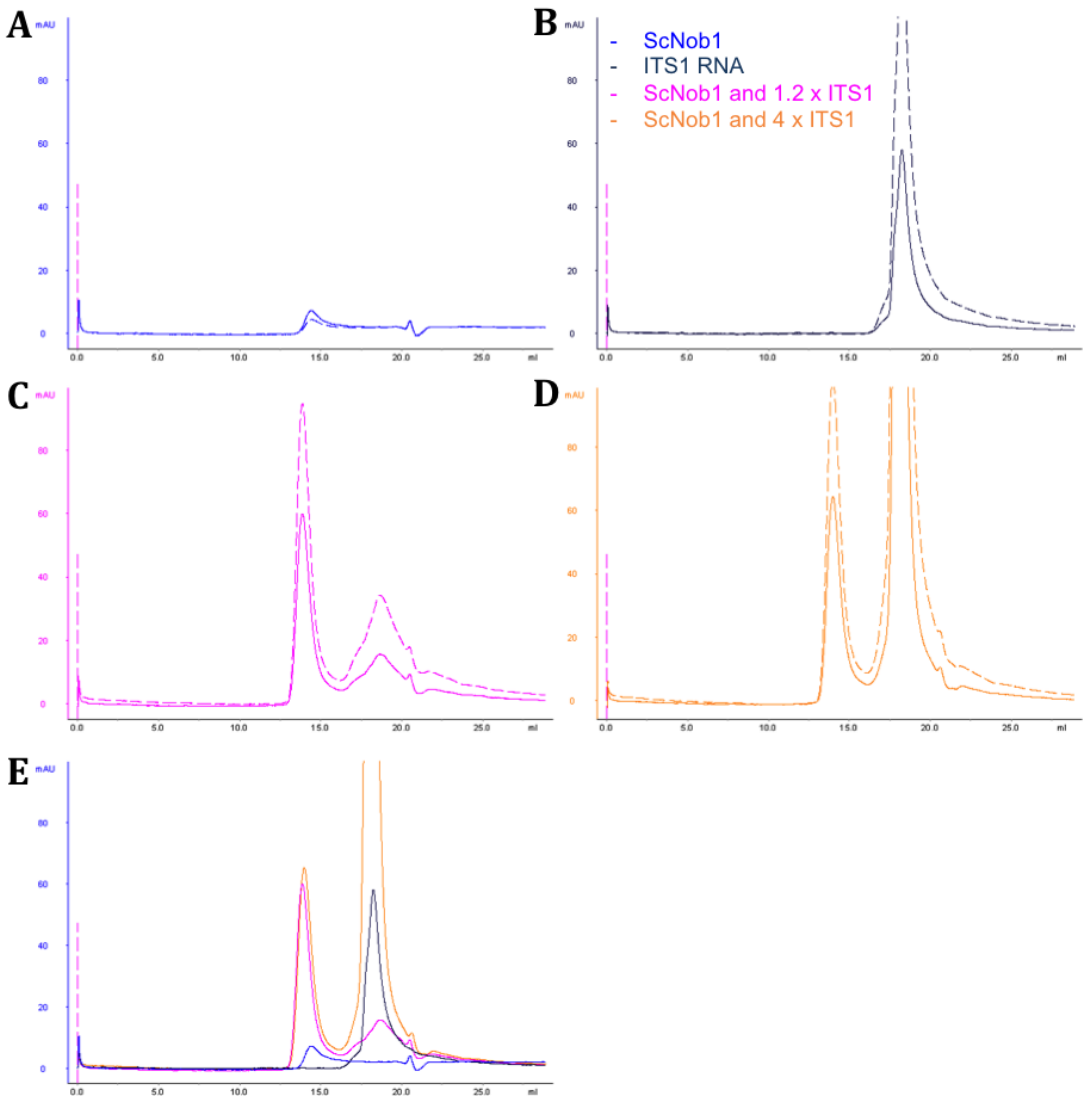


Figure 4.26: Semi-analytical SEC of ScNob1 and ITS1 RNA.
Absorbance at 280 nm is shown as a solid line whereas absorbance at 260 nm is shown as a dashed line. The key for the colours is shown on the right.

A) The SEC profile for ScNob1 only. A small single peak is observed.

B) The chromatogram for ITS1 RNA alone.

C) The SEC result when ScNob1 is mixed with ITS1 in a 1:1.2 ratio. The protein peak is seen to shift to an earlier elution volume and the absorbance at both wavelengths is significantly increased suggesting complex formation.

D) Results for ScNob1 and ITS1 mixed in a 1:4 ratio. Again, a shift and increase in absorbance are seen. However, the changes appear the same as in the previous experiment.

E) An overlay of the four previous results showing absorbance only at 280 nm. The increase in size and absorbance when the ITS1 RNA is added is clearly seen. The protein appears to be saturated at a 1.2 times excess of RNA as no significant further shifts are seen using a 4 times excess.

4.5 Nob1 can Cleave the D-site *in vitro*

Since ScNob1 can bind the D site *in vitro*, the next step was to check if it could perform the cleavage. DS5 was used for this purpose as it shows a large change in size when the cleavage occurs. It was found that the cleavage could happen *in vitro*. However, this process is manganese dependent (as expected) and is extremely inefficient (**Figure 4.27**). It also requires very high concentration of the metal in order to be observed.

The assay was repeated with a supplementary methylation step. The RNA was first incubated with purified Dim1 and *S*-Adenosyl methionine prior to incubation with ScNob1. However, no improvement in D site cleavage was observed (data not shown).

Finally, to confirm that the observed RNA is indeed the product of cleavage mediated by ScNob1 and not a contaminant, the nuclease dead mutant (D15N) protein was used in the assay. Due to purification issues, the protein was not completely pure and RNase contamination was seen. Nevertheless, the cleavage product was not present. **Figure 4.28** shows that the mutant is able to bind the RNA but does not appear to cleave it. The results are not conclusive due to the RNA digestion observed in the D15N mutant lanes.

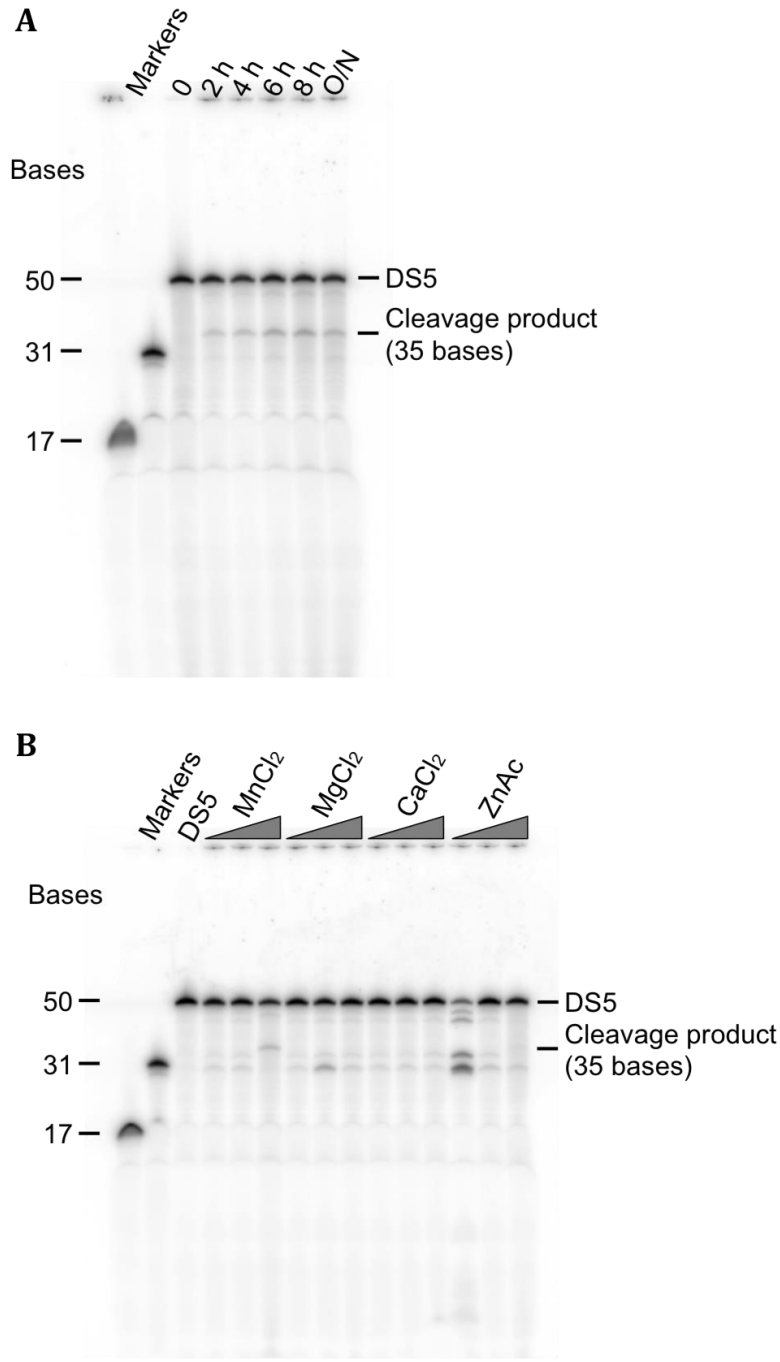


Figure 4.27: Cleavage assays using ScNob1 and DS5 RNA.

A) Time trial using manganese as the metal (5 mM). 0.2 µM RNA was incubated with 5 µM ScNob1. A cleavage product can be seen to appear after 2 hours of incubation at 37˚C. The amount of product increases at 6-8 hours. Overnight incubation, however, does not give an increased amount. Thus, a 6-hour incubation was chosen for the next experiment.

B) Cleavage assay using different metals at 50, 500, or 5000 µM. The RNA and protein concentrations are the same as before. The cleavage product is only seen when 5000 µM manganese is added to the reaction.

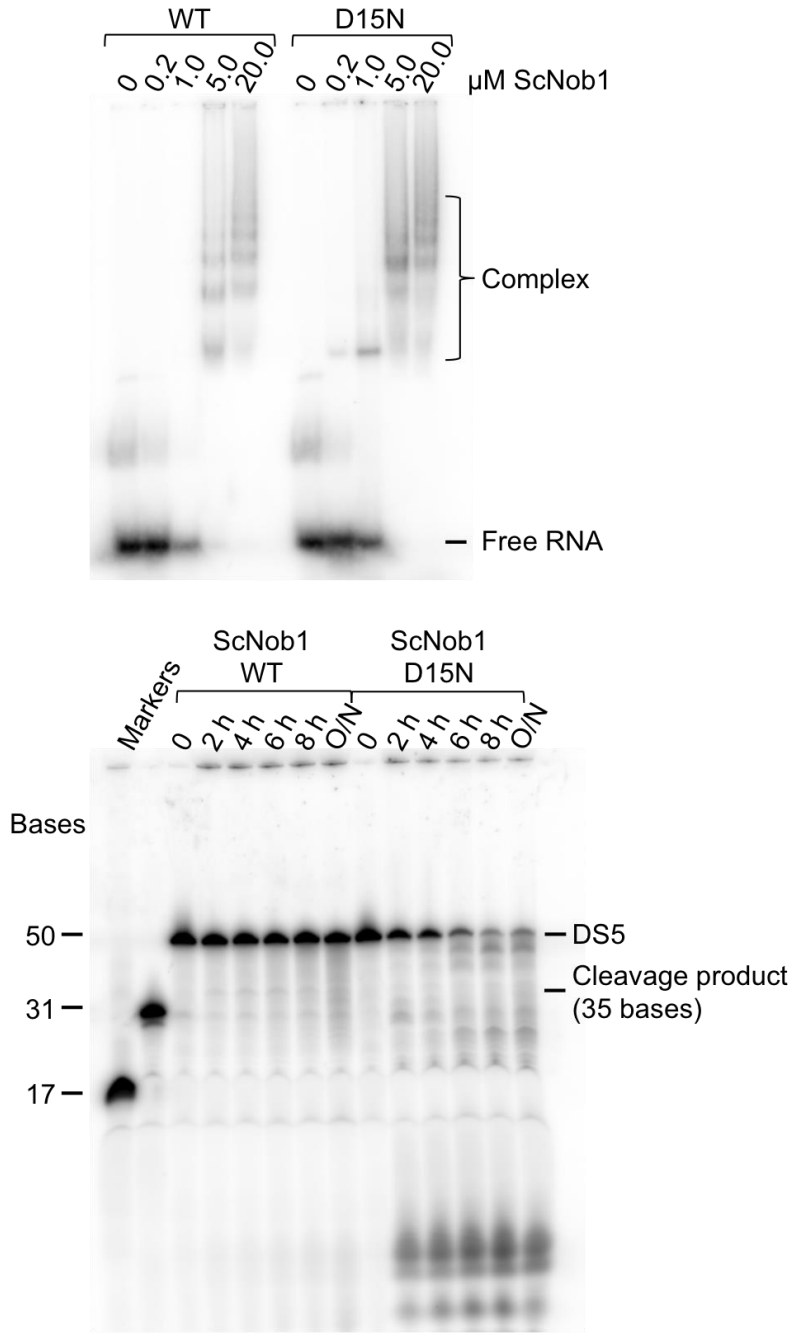


Figure 4.28: EMSA (top) and cleavage assay (bottom) comparing the WT ScNob1 and the D15N mutant.

The EMSA, in the top panel, shows that both proteins are able to bind to the DS5 RNA. Complex formation and higher order structures are seen earlier with the D15N mutant. This could mean that the mutant binds more strongly. Below, the cleavage assay using 5 μM of each protein is shown. A very small amount of product is seen using the WT protein. The D15N mutant shows various digestion products, especially at longer incubations. This suggests that there is RNase contamination present. However, the 35-base cleavage product cannot be seen at all. This could be obscured by the RNase digestion but is also likely to not be there at all, as this mutant was shown before to have no nuclease activity (Lebaron et al, 2012).

4.6 Crystallisation Trials using Protein-RNA Mixtures

As with the protein alone, crystallisation screens were performed with mixtures of ScNob1 and DS1, DS6, or ITS1. However, no hits were observed. Mixing the protein with any of the RNAs caused precipitation that disappeared upon incubation at room temperature. Since ITS1 appears to bind more strongly than D site RNAs (only a 1.2 times excess is needed to saturate the complex), this RNA was also used with ScNob1ΔL4. However, when the two components were mixed, precipitation was observed and it could not be dissipated by incubations at higher temperatures or in higher salt.

In the future, further screens should be performed using the loop-out mutants of ScNob1 with different RNAs, particularly with ScNob1ΔL2 and ΔL3 as they can definitely form crystals. Shorter ITS1 RNAs that do not form a hairpin might prevent precipitation. The recognition site in this RNA needs to be identified to define the minimal binding requirements. Also, tests with the mutant proteins and D site RNAs should be performed.

4.7 Purification of ScNob1-D site RNA complex

For further crystallisation trials, the ScNob1 and DS1 RNA were co-purified as a complex. This required large scale *in vitro* transcription followed by RNA purification. Initially, SEC was followed by anion exchange. **Figure 4.29** shows the results of this purification. Multiple peaks were obtained in the anion exchange. To determine which peak was best for binding to ScNob1, semi-analytical SEC was performed. It was found that peak 3 was best (**Figure 2.30**) and this was chosen for the next step.

The ScNob1 protein was partially purified (affinity chromatography followed by cation exchange). Then, the transcribed and purified peak 3 RNA was added and a SEC step was performed. **Figure 4.31** shows the results of the co-purification.

The obtained complex was used in crystallisation screens. However, no hits were obtained. The same experiment should be repeated with the ITS1 RNA in the future.

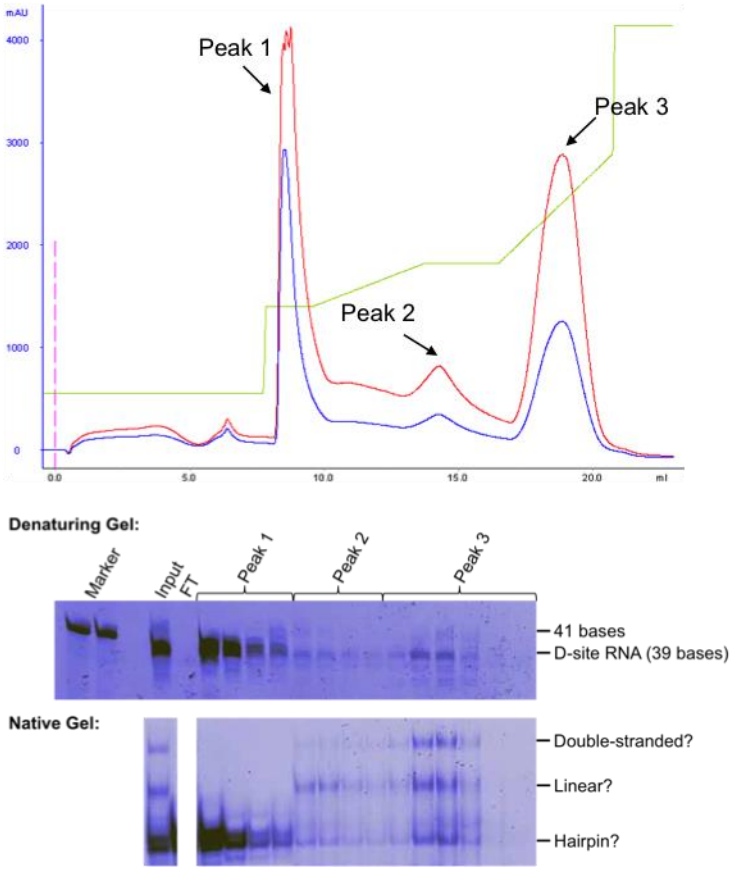


Figure 4.29: Mini Q anion exchange purification of *in vitro* transcribed DS1.
A single peak was obtained in SEC, which was then injected onto the anion exchange column. The gradient included 3 steps, which were optimised prior to the large scale run shown here. The first step consisted of a wash with 35% buffer B (buffer A: 20 mM HEPES pH 7.5 and 5 mM $MgCl_2$, buffer B: 20 mM HEPES pH 7.5, 5 mM $MgCl_2$ and 1000 mM KCl). Then, a gradient from 35 to 45% over 4 column volumes was used and the column was washed at 45%. Finally, a gradient from 45 to 70% buffer B eluted the last peak. The three peaks were analysed on denaturing and non-denaturing gels. They all contained RNA of the right size (denaturing gel). However, the RNA showed different conformations (native gel).

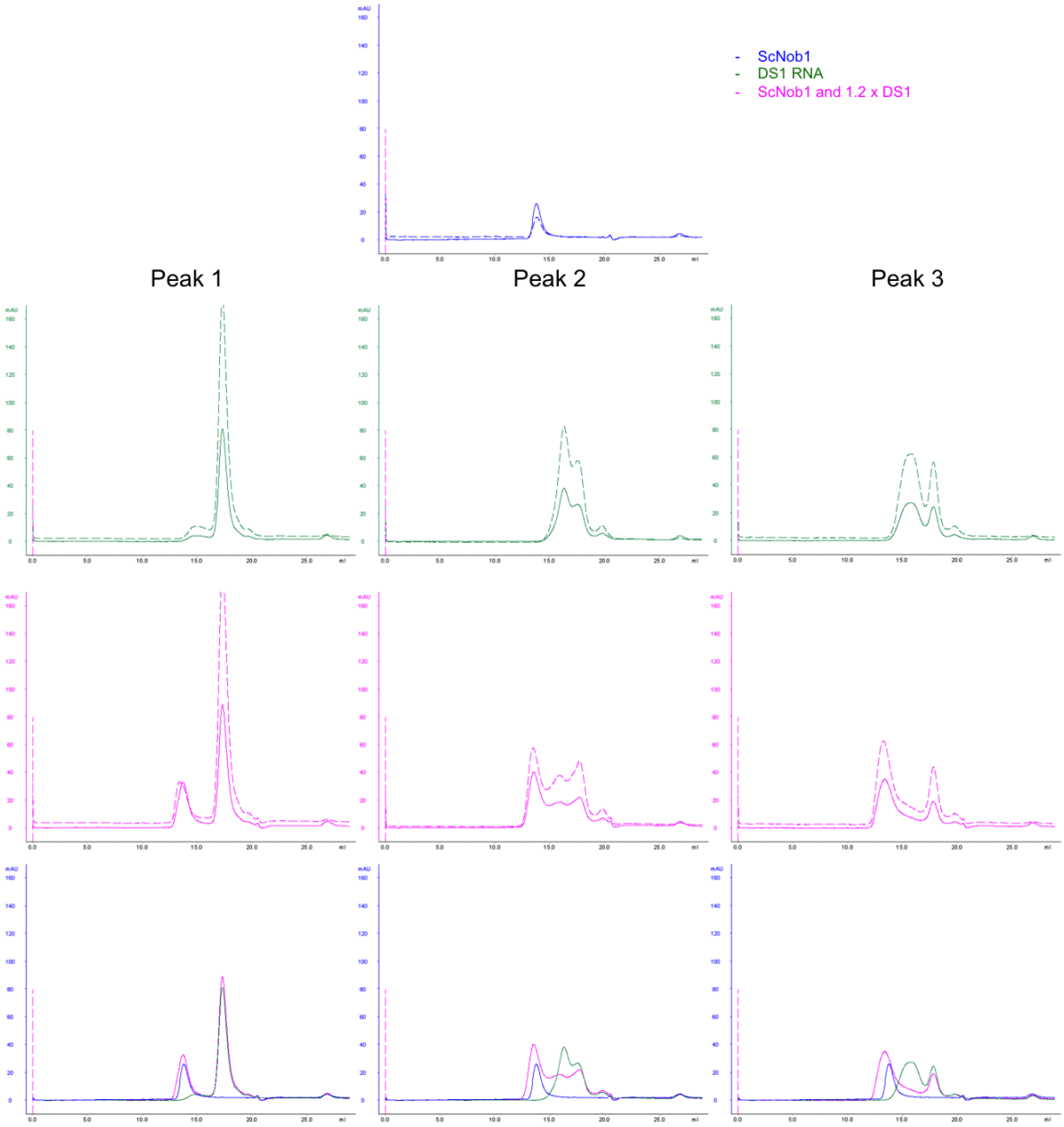


Figure 4.30: Semi-analytical SEC results using ScNob1 with each of the DS1 peaks obtained from the Mini Q column.

The legend is shown in the top right corner. As before, the absorbance at 260 nm is shown as a dashed line while absorbance at 280 nm is shown as a solid line. Peak 1 showed the smallest species in SEC. However, it failed to show significant binding to ScNob1. Peak 2 showed the presence of two species. The larger one was depleted upon addition of ScNob1 suggesting an interaction. Peak 3 had the most complex elution profile. However, it also showed the largest shift upon ScNob1 addition and was therefore used for the large-scale complex formation.

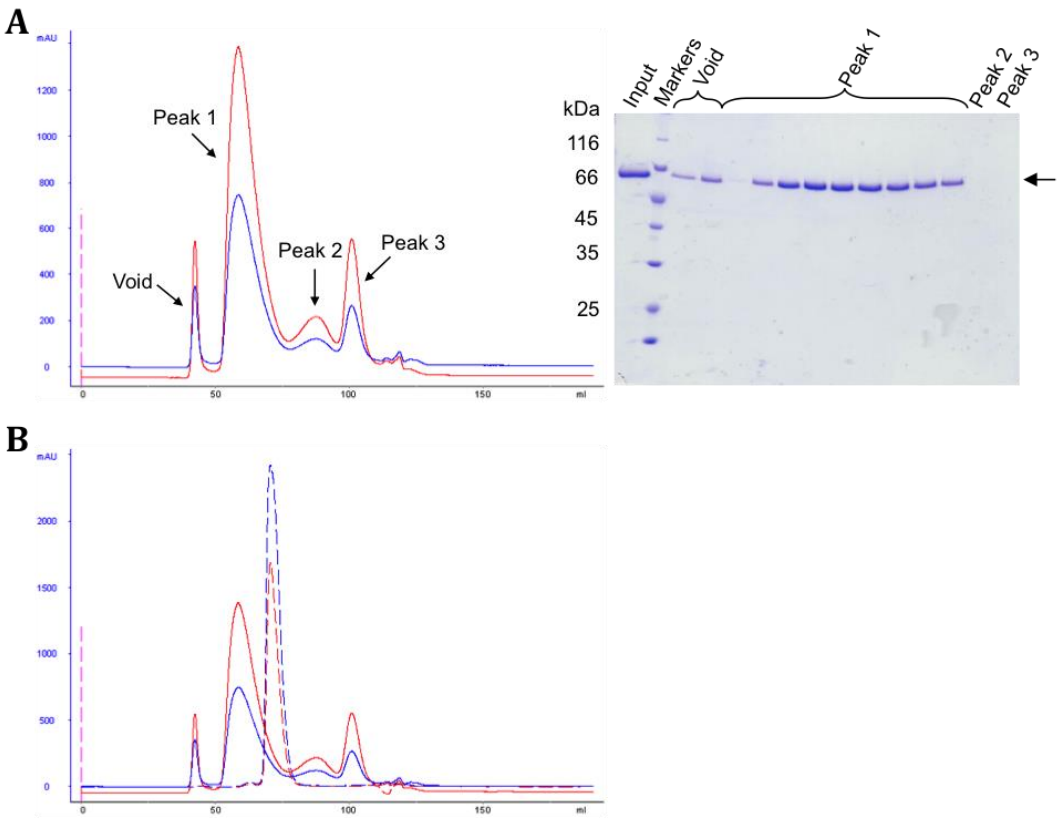


Figure 4.31: Co-purification SEC results for ScNob1 and DS1.
Absorbance at 280 nm is shown in blue and at 260 nm in red.

A) The elution profile for the complex. Three peaks were seen. The void and the first peak contained the ScNob1 protein. Thus, the excess RNA was likely in the other two peaks. The absorbance at 260 nm in the first peak suggests that a complex was formed.

B) An overlay of the complex elution profile (solid lines) and the profile for ScNob1 on its own (dashed lines). The fractions in peak 1 that did not overlap the free protein peak were pooled and used for crystallisation screens.

4.8 Discussion

4.8.1 ScNob1 binds to D site and ITS1 but not helix 40 RNAs

EMSAs, semi-analytical SEC, and ITC have shown that the recombinant ScNob1 is able to bind certain D site and ITS1 sequences. However, it does not show any interaction with the H40 RNA that was previously found in *in vivo* cross-linking studies (Granneman et al, 2010). Most of the experiments involving the H40 RNA used the short hairpin stabilised by extra bases. Thus, it is possible that, if ScNob1 only interacts with single-stranded RNA at this site, this caused the lack of

interaction. However, a specific interaction with H40 is unlikely as the binding mode would not be the same as for the D site (no AGGA sequence is present in H40). It would also be distinct from binding seen at ITS1 as this RNA fragment likely forms a hairpin but is still able to interact with the protein.

This is supported by the EMSAs, which were performed with wild-type RNA fragments (no extra bases were included), as these also showed no significant binding. The same was observed in SEC with the intermediate H40 RNA fragment, which contained further sequences that could have influenced binding. Nevertheless, all of the H40 sequences tested would have formed stable hairpins. If Nob1 requires single-stranded regions (like the ends of the ITS1 fragment), no interaction would be observed with the H40 fragments. Otherwise, H40 could have been cross-linked to Nob1 due to proximity and not due to complex formation, or other factors might be needed to stabilise the interaction.

The secondary structures of the H40 fragments were confirmed using primer extension assays. However, as seen with the intermediate fragment in SEC, multiple conformations still exist. If two separate RNA molecules were to hybridise, the loop structures would still appear in the same places in primer extension, as it is not sensitive to changes that do not affect secondary structure. The amount of alternative structures observed is exacerbated because the SEC was performed at 4˚C while the primer extension was at room temperature. Finally, the concentration of RNA used might have a large effect on the dimerization of the RNA molecules. At higher concentrations (as in SEC), there is a higher chance of dimer formation.

The long H40 RNA was not used because data analysis would be very complex with such a complex fragment and its full structure was not determined. A second primer could have been used to determine the structure as the 5′ end. However, this was not performed, as there was no evidence of Nob1-H40 interaction found. The lack of binding was supported by the primer extension protection experiment with all three H40 RNAs. The pattern of 1M7 modification did not change when ScNob1 was included in the reactions suggesting that there were no interactions, even with the longest H40 fragment.

Conversely, most D site fragments used in the different experiments were able to interact with ScNob1. The AGGA sequence was found to be needed for the

interaction. Changing the guanines into adenines and vice versa made the interaction weaker. However, changing the last adenine to a uracil had no effect. The reason for this remains unclear. It could be that the binding site is able to accommodate a change into a smaller base or that the effect of the mutation on the secondary structure of the RNA is smaller in this case. Further experiments would be needed.

Either way, it appears that Dim2 and Nob1 are likely to compete for this site. It is possible that Dim2 prevents Nob1 binding at early stages and cleaving before all the other maturation steps have been completed.

Finally, an ITS1 fragment was tested in the ITC and SEC. This fragment was chosen as some previous studies have shown that it can interact with Nob1 (Lamanna & Karbstein, 2009). However, the data suggesting this interaction were not conclusive as only a single base showed changes in modification when Nob1 was added. The ITC performed here with a short RNA fragment showed that it binds but the curve was too complex to fit a one-site binding model. EMSAs and further ITC experiments should be performed to further explore the recognition sequence within the ITS1 RNA.

ITC competition experiments should also be performed to determine if the D site and ITS1 RNAs could bind to the protein at the same time. For this purpose, ScNob1 would initially be saturated with one RNA. Then, the other one would be titrated in. If the K_d remains the same as without the competing RNA, the two fragments bind independently. A lower K_d would suggest co-operative binding while a higher one would show competition for a binding site.

It is possible that the D site is bound by the PIN domain while the ITS1 is bound by the zinc ribbon domain. New CRAC data in growing cells support that ScNob1 binds to both, the D site and ITS1. The H40 site is also still present in these results (personal communications with Prof. Tollervey) adding more complexity. The SEC results using ITS1 show complex saturation at a 1.2 times excess of RNA. With the D site, a four times excess is needed suggesting that the ITS1 binds more strongly or that fewer sites are present. Since the ITS1 fragment does not contain an AGGA sequence and has a significantly different structure (tight hairpin), it is likely that different binding mechanisms and/or sites are involved.

In the future, the loop-out mutants of ScNob1 should also be tested to see if they bind the same RNAs with the same affinity. The experiments could also be used with CtNob1 to see if the binding sites are conserved between the two organisms.

4.8.2 Large complexes are formed with the D site RNA and ScNob1

Multiple results suggest that ScNob1 forms large, heterogeneous complexes when the D site is added. The EMSAs showed multiple bands, which suggest that multiple proteins/RNAs were able to assemble together. The bands were well-defined (not a smear of variable size) suggesting a distinct, modular behaviour. The amount of these higher weight bands correlates to protein concentrations used, further suggesting that multiple proteins come together to make these. This could be biologically relevant if a multimer of the protein is required to perform the cleavage. This is unlikely, as the previous EM studies do not support a multimer binding to the pre-40S (Strunk et al, 2011). However, it could potentially provide another level of control for the final step of maturation.

In the EM, addition of D site RNA caused the formation of long structures. Analysis of the structures was not undertaken due to time constraints. However, these are unlikely to form inside the cell, as there would not be enough space in the busy environment around the premature ribosomes. The local concentrations would be too low and the rest of the ribosomal structure would prevent the particles coming close together.

In SEC-MALS, ScNob1 and DS1 were seen to elute in the void volume supporting the formation of a mega-Dalton species. This was not seen in the semi-analytical SEC performed locally, likely because the sample was fresh and kept at 4˚C. For the SEC-MALS, the mixture of protein and RNA had to be shipped and was only tested a day later. Also, the SEC-MALS was run at room temperature while our SEC was at 4˚C.

The molar ratio obtained from ITC suggested that four molecules of RNA were able to bind to one molecule of protein. This, again, implies that complex structures can be formed when the two particles come together. As before, the biological significance for this remains unclear. Although the modular behaviour could form a part of the control mechanism for the cleavage, the complexes seen in

EM would not be able to form inside the cell. Overall, further studies are needed to determine if the cleavage efficiency is affected when the multimerisation is prevented. Also, tests to see whether the loop-out mutants behave in the same manner should be performed.

Finally, semi-analytical SEC with the human DS1 showed elution of a spectrum of species. They ranged in elution volume from the void to the normal ScNob1 volume. This was not seen with the yeast RNA sequence although the two only differ in the 3′ end. The effect could be partially explained by the fact that the human sequence shows a higher propensity for secondary structure formation (ΔG = -2.60 kcal/mol vs. -0.30 kcal/mol predicted by mFold - http://mfold.rna.albany.edu) and dimerization than the yeast. Thus, the RNA-RNA interactions could be causing different species to form.

4.8.3 Binding is likely affected by secondary structures

Initially, it appeared that DS0 could not bind to ScNob1 (while DS1 RNA could) because of the hairpin formation. However, once the binding site was identified with ITC, it became apparent that DS0 contains a point mutation in this site (GGGA instead of AGGA). Thus, the lack of binding could have been caused by the mutation and not the secondary structure. DS12 was designed to distinguish between the two possibilities. It contained the same mutation as DS0 but lacked the 3′ end, which allowed hairpin formation. This RNA was seen to give a very small heat change in ITC suggesting that no significant interactions were occurring. DS0 shows no heat change at all suggesting that both the structure and sequence are causing the effect.

The RNAs used for ITC were checked with mFold and showed no predicted secondary structures. Nevertheless, the native gel of the RNAs showed that most of the mutants that impaired ScNob1 binding (DS6.1-DS6.5: AGGA to GGGA, AAGA, AGAA, AAAA, or AGGG) showed an increased amount of secondary structures. When a mutation that did not cause secondary structure formation was introduced (DS6.5.2), the binding was rescued. However, this mutation only included the last nucleotide (AGGA to AGGU). Thus, it is possible that this base is less important than the middle ones. Also, the adenine-to-uracil substitution reduces the size of the

base (unlike the adenine-to-guanine experiment) and could, therefore, be more tolerated. Additionally, DS6.4, the double mutant (AAAA sequence), did not have additional structures but showed less binding than the single mutants. Thus, the sequence is definitely important. Further point mutations, particularly adenine/guanine to uracil, need to be tested.

The native gel also showed that the RNA concentration had a large effect on the structure. When the same total amount of DS0 was separated on the gel but in a smaller volume (i.e. higher concentration), structures previously not seen were observed. The concentration used in ITC was relatively low (10 μM). Thus, it is less likely that the structures observed on the gel actually occurred in ITC. Again, this means that the effect of RNA structure on ScNob1 binding could not be fully determined.

In contrast to the DS0 hairpin, the ITS1 hairpin was able to bind to ScNob1 in the ITC. However, as mentioned earlier, it is possible that the two RNAs interact with different domains of the protein. Thus, their interactions are not directly comparable.

In the future, longer D site RNAs with further mutations could be tested in ITC. Their structures could be probed with primer extension at the concentrations that they would be used at in the experiments. The DS6-based RNAs used here were too short for primer extension. However, others (like DS5) could be utilised. *In vivo*, the secondary structure of the D site RNA remains in question, although there is some evidence to suggest that the predicted hairpin is not formed and the D site is single-stranded (Lamanna & Karbstein, 2009).

4.8.4 *In vitro* cleavage is very inefficient

The recombinant ScNob1 protein was able to perform the D site cleavage. As previously observed with longer substrates (Lebaron et al, 2012), this cleavage was very inefficient and required large amounts of manganese. However, this result was obtained when Nob1 was taken out of the context of the ribosome. It is possible that some other factors are needed for the cleavage to occur efficiently. For example, the RNA was pre-incubated with Dim1, potentially methylating it. However, the

methylation state was not confirmed and the pre-incubation did not help with the cleavage.

A methylation assay would need to be developed in the future to see if the Dim1 protein is active *in vitro*. Other factors, such as Rio2, might also be required for Nob1 to perform its function. Additionally, the high levels of manganese required *in vitro* suggest that a cellular mechanism must exist that compensates for the low levels inside the cell. Whatever this mechanism is, it might be required for cleavage. For example, Nob1 might have to be phosphorylated in order to bind manganese or cleave RNA.

Finally, a D site cleavage product mimic was used in ITC (DS11). This RNA was able to bind to the ScNob1 protein as well as fragments containing the 3′ end. Thus, additional factors may be needed to release the substrate. These might also be needed for the cleavage to occur efficiently. Since Nob1 must cleave with high fidelity in order to maintain the ribosomal function, it is likely that all these factors are there to make sure that the cleavage only occurs in particles that are otherwise correctly processed and at the right time.

The cleavage assay was then performed with a nuclease dead mutant protein (D15N) to confirm that the observed band was created by ScNob1 activity. The mutant protein, however, was difficult to purify and contained RNase contamination. Thus, it remains unclear whether the predicted cleavage product seen on the gel is definitely created by Nob1, although it is the correct size and only appears when Nob1 is included in the reaction suggesting that it is likely to be the product. In the future, if clean ScNob1 D15N can be obtained, the EMSA and cleavage assay should be repeated to show that the binding to the D site is not affected while the cleavage is inhibited.

Mutant RNAs that do not bind to the wild-type protein (e.g. AAAA sequence instead of AGGA) should also be tested in the cleavage assay to see if binding is required for cleavage. The existing loop-out mutants could also be used, to test if the loop is necessary for nuclease activity. ITC would be used to see if the same deletions affect the protein-RNA interactions.

4.8.5 Crystallisation trials were unsuccessful

All crystallisation trials with ScNob1 and different RNAs were unsuccessful. The K_d for the D site RNAs appears to be around 1.0 µM. Thus, when the excess of RNA is removed or the sample is diluted, the complex can separate again. Also, as seen in some of the experiments, the RNA and protein complex can form larger structures at high concentrations. This might be impeding the formation of crystals, as it would introduce heterogeneity.

Finally, if the RNAs interact with only one domain, the two domains could still be too flexible with respect to each other. More likely, the flexibility of the loop between the domains is preventing crystal formation (as the loop-out mutants were seen to form crystals). Therefore, in the future, the loop-out mutants in complex with RNA should be tested.

Additionally, separating the two domains might prove to be successful in the future, although initial attempts did not work. If this is the case, they should be used with the D site and ITS1 RNA fragments separately to see which domain interacts with which RNA fragment. The other option is to make a D site RNA that is connected (through an RNA sequence with no secondary structures) to the ITS1 RNA. Thus, if the two fragments interact with different domains, this long RNA could 'lock' the protein in a single conformation. Different lengths of connecting sequences could be used to see what gives the strongest interaction.

4.9 Concluding Remarks

The ScNob1 D site RNA binding site was successfully identified. Co-crystallisation trials were unsuccessful to date and the molar ratio remains in question. However, these results are a good start for further experiments and RNA binding partner optimisation. The ITS1 fragment also looks like a promising RNA species that could be used in the future.

The D site binding sequence appears to be only four nucleotides long. Thus, other mechanisms would have to be involved in positioning the protein correctly and

ensuring the cleavage occurs in the right place. The binding to ITS1 could potentially be more specific and be a part of this mechanism.

5 CHARACTERISATION OF DIM2 AND ITS INTERACTIONS

5.1 Introduction

Dim2 is a highly conserved RNA binding protein that, in eukaryotes, consists of a degenerated (KH1) and a canonical (KH2) KH domain. It is required for the A_2 and the D site cleavages and, thus, localises to both the nucleus and the cytoplasm (Vanrobays et al, 2004). It has been shown to interact with Nob1 through *in vitro* co-precipitation (Woolls et al, 2011) and yeast-2-hybrid assays (Tone & Toh, 2002). The interaction appears to be regulatory, as Dim2 does not exhibit any enzymatic activity.

The regulatory role of Dim2 is also important for the dimethylation reaction performed by Dim1. These two proteins have also been shown to interact through co-precipitation in cell extracts (Vanrobays et al, 2004). Dim2 is also involved in the pre-ribosome export from the nucleus (Vanrobays et al, 2008). As with Nob1, a single mutation in Dim2 (G207A) can prevent 20S processing (Vanrobays et al, 2008). This mutation is in the RNA binding loop of the conserved KH domain and is predicted to prevent the protein from binding to the pre-ribosome.

An RNA recognition site for Dim2 can be predicted from an X-ray crystal structure of the archaeal protein from *P. horikoshii* (Jia et al, 2010), as shown in **Figure 1.8** in the first chapter. This site appears to be very close to the D site, supporting the protein's role in the final stages of maturation. However, unlike the eukaryotic protein, the archaeal protein contains two canonical KH domains. Thus, the RNA recognition site in eukaryotes requires further exploration to determine whether it is the same as in archaea. This chapter examines the interaction between recombinant Nob1 and Dim2, and the D site RNA in the eukaryotic system, providing insight into how the final D site cleavage might be regulated.

5.2 ScDim2 Modelling Based on the Archaeal Structure

Initially, a model for the yeast protein was created using crystal structures of the archaeal Dim2. A multiple sequence alignment using Dim2 from eukaryotes and archaea was entered into Phyre, Phyre 2, and Swiss-Model. All three programs used archaeal structures (PDB IDs: 2E3U, 1TUA, and 3AEV respectively) to create a model for the yeast protein based on the predicted similarity to the solved structures.

Both, Phyre and Swiss-Model, used the *P. horikoshii* structure (without and with RNA respectively). Phyre 2 used a structure of a protein of unknown function from *Aeropyrum pernix*. **Figure 5.1** shows a sequence alignment of Dim2 from the two archaeal species and human and yeast. All three generated models were very similar showing an RMSD of 1.1 (Phyre to Phyre 2), 0.5 (Phyre to Swiss-Model), or 1.3 Å (Phyre 2 to Swiss-Model). Thus, one model, generated by Phyre, was chosen for further analysis.

Phyre generated a model of residues 95-266 from the yeast protein (274 amino acids in length). As shown in **Figure 5.2**, the general fold of the modelled ScDim2 is very similar to that of the *P. horikoshii* protein. The locations of the two GxxG loops are marked on the figure. In **Figure 5.2B** only these loops from the archaeal structure and the generated model are shown. The conserved loop in the KH2 domain fits well into the RNA. However, the degenerated sequence in ScDim2 KH1 shows large amino acids that do not interact with the RNA in the ideal fashion.

Further model validation was undertaken, as presented in **Figure 5.3**. The geometry of the side chains was analysed (**Figure 5.3A**). Multiple outliers were found, especially in KH1. However, when these were mapped onto the model (**Figure 5.3B**) it was found that most of them were on the protein surface. Thus, their orientations could be readily adapted without disrupting the internal arrangement. Finally, a Ramachandran plot was also created (**Figure 5.3C**) to analyse backbone geometry. Only three residues were found to be in non-allowed regions.

Overall, the model appears to be suitable, particularly for the KH2 domain (residues 195-247). The majority of KH1 (99-151) residues also appear to occupy plausible orientations. The largest differences are seen towards the end of this domain and in the RNA binding loop of KH1 (NSWT instead of GxxG). In the RNA

binding loop, while the side chain geometry is allowed in the actual loop, the preceding arginine (110) shows side chain angles that are unlikely to occur in nature and invades the predicted RNA binding site. Also, the backbone geometry in this location is inconsistent with naturally occurring angles, as asparagine 111 (the first residue of the loop) is one of the Ramachandran plot outliers. Thus, a eukaryotic structure is highly desirable to obtain structural insight into the fold of the Dim2 protein, particularly in the degenerated domain. This, in turn, might lead to insights into its function, as it appears to be unable to bind to RNA in the same fashion as the archaeal Dim2 KH1 domain.

The inability of the eukaryotic KH1 domain to bind rRNA is supported by the lack of conservation of the predicted rRNA binding site between eukaryotes and archaea.

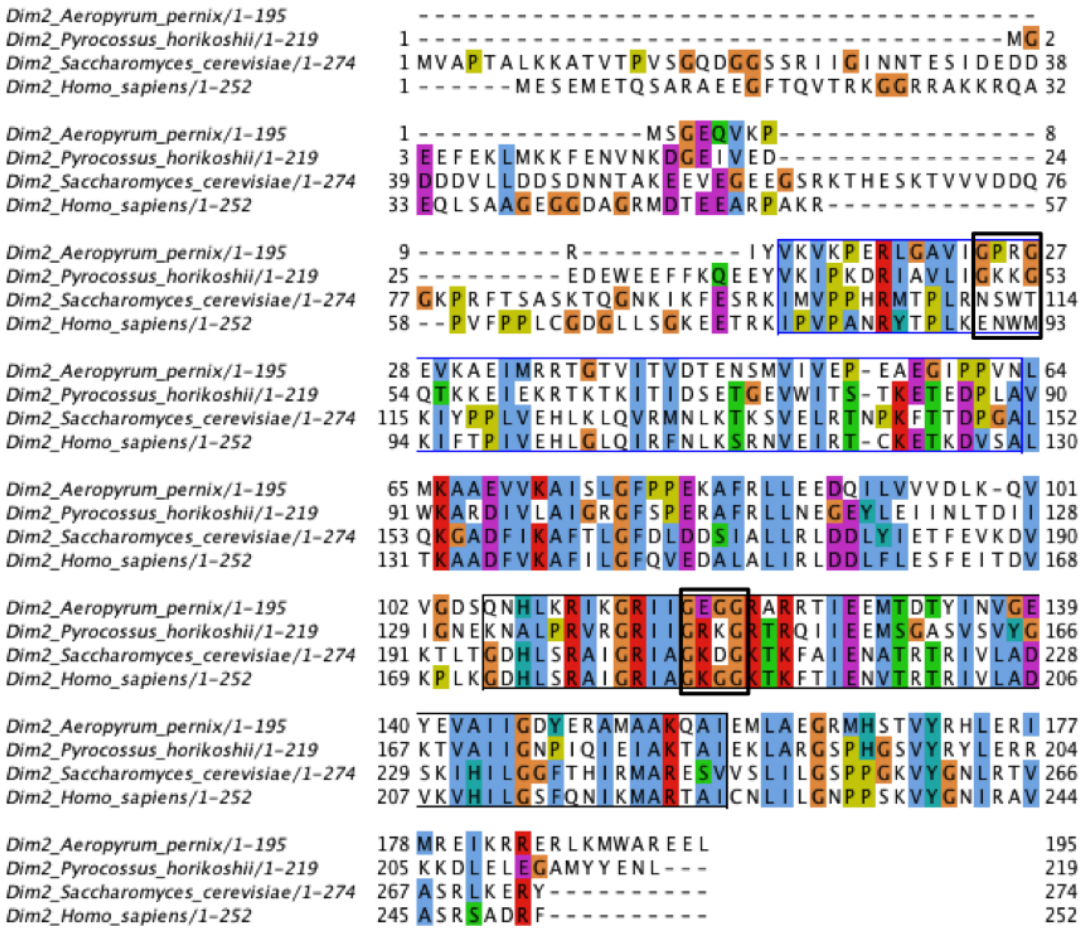


Figure 5.1: Dim2 sequence alignment using two archaeal sequences and the human and yeast proteins.

The degenerated KH1 domain is marked in a thin blue box while the conserved KH2 domain is in the thin black box. The RNA binding loops (GxxG) are shown in the thick black boxes in each domain. The KH1 loop in archaea shows the GxxG motif while different sequences are seen in eukaryotes.

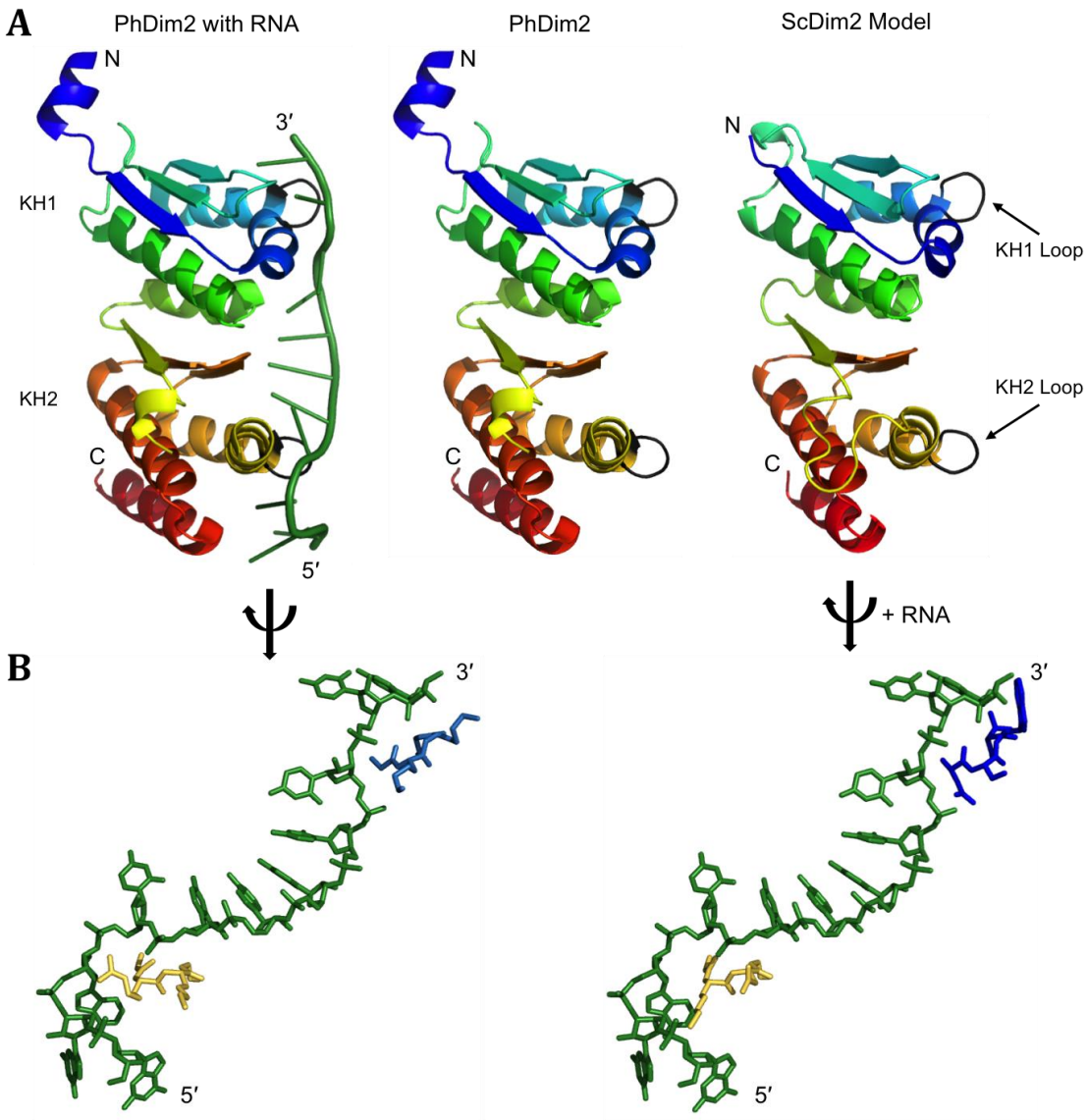


Figure 5.2: Structural comparison of the *P. horikoshii* (3AEV) structure (PhDim2) to the ScDim2 model generated by Phyre.

A) PhDim2 with and without RNA is shown next to the predicted ScDim2 structure. The proteins are coloured from the N-terminus (blue) to the C-terminus (red). The RNA binding loops in each domain are shown in black and indicted with arrows on the right. The RNA is shown in green in the first image.

B) A view of the RNA binding loops (blue and yellow) and RNA (green) in the PhDim2 solved structure (left) and modelled ScDim2 (right). The RNA and protein fragments are shown as sticks. The KH1 loop (blue) in the modelled structure shows larger amino acids that do not fit in the RNA groove. The conserved KH2 loop (yellow), in both cases, shows small residues that fit well into the RNA. The two glycines (GxxG) allow the short loops to form tight turns in order to be incorporated in this manner.

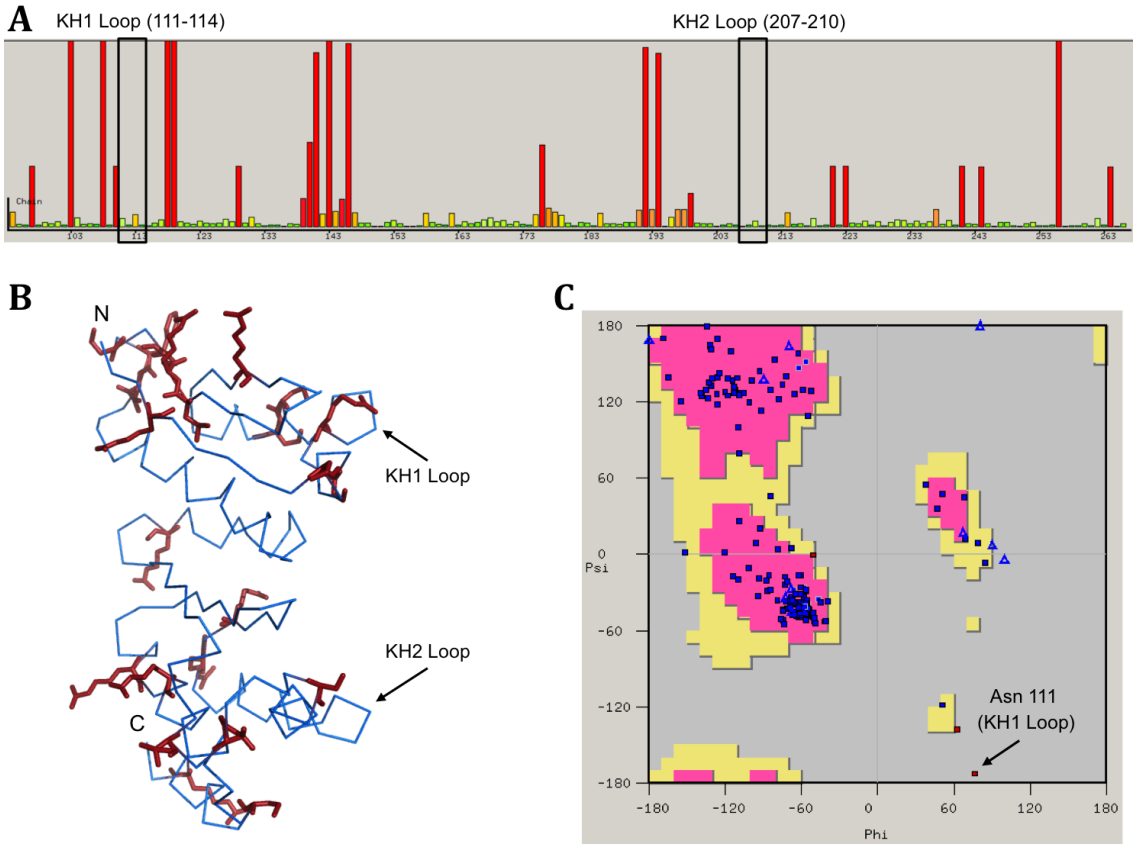


Figure 5.3: Model validation using side chain and backbone geometry.

A) Geometry analysis of the amino acid side chains in the model. The red bars represent residues that are in unsuitable orientations. The taller the bar, the more unlikely the orientation for that amino acid is imposed. More of the tall red bars are located in the N-terminal KH-like domain. Overall, 23 out of 266 residues are red.

B) The geometric outliers are shown with their side chains as red sticks while the rest of the protein backbone is shown as a blue ribbon. The majority of the outliers are pointing out onto the protein surface. Thus, their orientations could be changed without disrupting the overall arrangement.

C) Ramachandran plot for the predicted model. Three residues were found to be in the disallowed regions. Nine were allowed while the rest (159) was found in the preferred regions.

5.3 Cloning and Expression Trials

Multiple constructs for Dim2 were created in order to study its function. These are summarised in **Table 5.1**, which also summarises the results of the expression trials performed with these constructs. The tests for the full-length yeast protein were repeated and can be seen in **Figure 5.4**.

Table 5.1: Summary of the expression results for different constructs created.
The proteins were expressed as fusions with hexahistidine (His) or glutathione S-transferase (GST) tags. The first two constructs were present prior to the start of the project.

Insert	Organism	Plasmid	Tag	Comments
Dim2 FL	*S. cerevisiae*	pEC-KHC, pEC-KGC	GST or His	Expressed, Largely soluble
Dim2ΔN91	*S. cerevisiae*	pEC-KGC	GST	Expressed, Some soluble
Dim2ΔN91	*S. cerevisiae*	pEC-KHC	His	Expressed, Soluble
Dim2ΔN51	*S. cerevisiae*	pEC-KHC	His	Expressed, Soluble
Dim2ΔN61	*S. cerevisiae*	pEC-KHC	His	Expressed, Soluble
Dim2ΔN99	*S. cerevisiae*	pEC-KHC	His	Not much expressed, Insoluble
Dim2 FL	*S. cerevisiae*	pEC-SHT	His	Expressed, Soluble
Dim2 FL	*S. cerevisiae*	pEC-SGT	GST	No expression
Dim2ΔN31	*H. sapiens*	pEC-KHC	His	Expressed, Soluble
Dim2ΔN42	*H. sapiens*	pEC-KHC	His	Expressed, Soluble

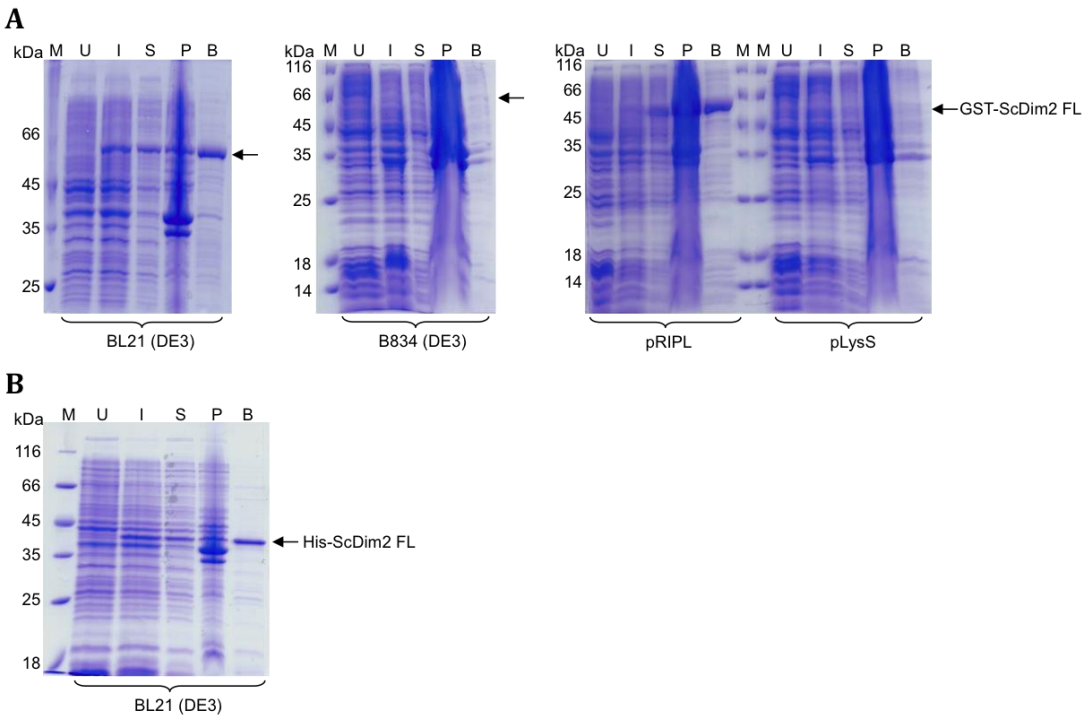


Figure 5.4: Expression trials of full-length (FL) ScDim2 constructs present prior to the start of this project.
The different lanes correspond to the different fractions: uninduced (U), induced (I), soluble (S), insoluble/pellet (P), and the nickel/GSH bead-bound (B) fraction. The molecular weight markers (M) are also shown on each gel. The band corresponding to the correct size is indicated with an arrow. The same labels are used in all following expression trials.

A) Expression of the full-length GST-tagged protein in different cell lines in 2xTY medium. The BL21 (DE3) and BL21 (DE3) pRIPL cell lines show significant soluble expression and enrichment in the bead-bound fraction. The BL21 (DE3) strain was chosen for further cultures.

B) Expression of His-tagged ScDim2 in BL21 (DE3) cells in the 2xTY medium. Less overexpression is seen than with the GST-tagged protein. However, the protein is still soluble and enriched on the beads.

Additional constructs were also made and tested for expression (**Table 5.1**). The initial truncated construct (ScDim2ΔN91) was GST-tagged. However, when the GST was removed with 3C protease, the protein precipitated out of solution. The presence of the large, soluble tag masked the insolubility of the construct. Thus, the majority of constructs made contained His-tagged protein in order to avoid these masked solubility issues. **Figure 5.5** shows the expression test results for a series of His-tagged yeast constructs. ScDim2ΔN91 was seen to be mostly soluble with both types of tag (the previously performed trials for the GST-tagged protein were not repeated).

The human GST-tagged Dim2 sequence was obtained from Dr. Nick Watkins (University of Newcastle) in the pGEX-6P-1 plasmid. The construct contained a 3C cleavage site and the full-length protein. Expression tests were performed and are shown in **Figure 5.6A**. Good, soluble expression was seen. As shown in **Table 5.1**, the protein was re-cloned into the pEC-KHC plasmid containing different N-terminal truncations. These two constructs also showed good, soluble expression (**Figure 5.6B**).

Finally, the full-length protein from *C. thermophilum* was ordered as a codon optimised plasmid. **Figure 5.7** shows that no soluble expression of this protein was obtained. Thus it was not used any further.

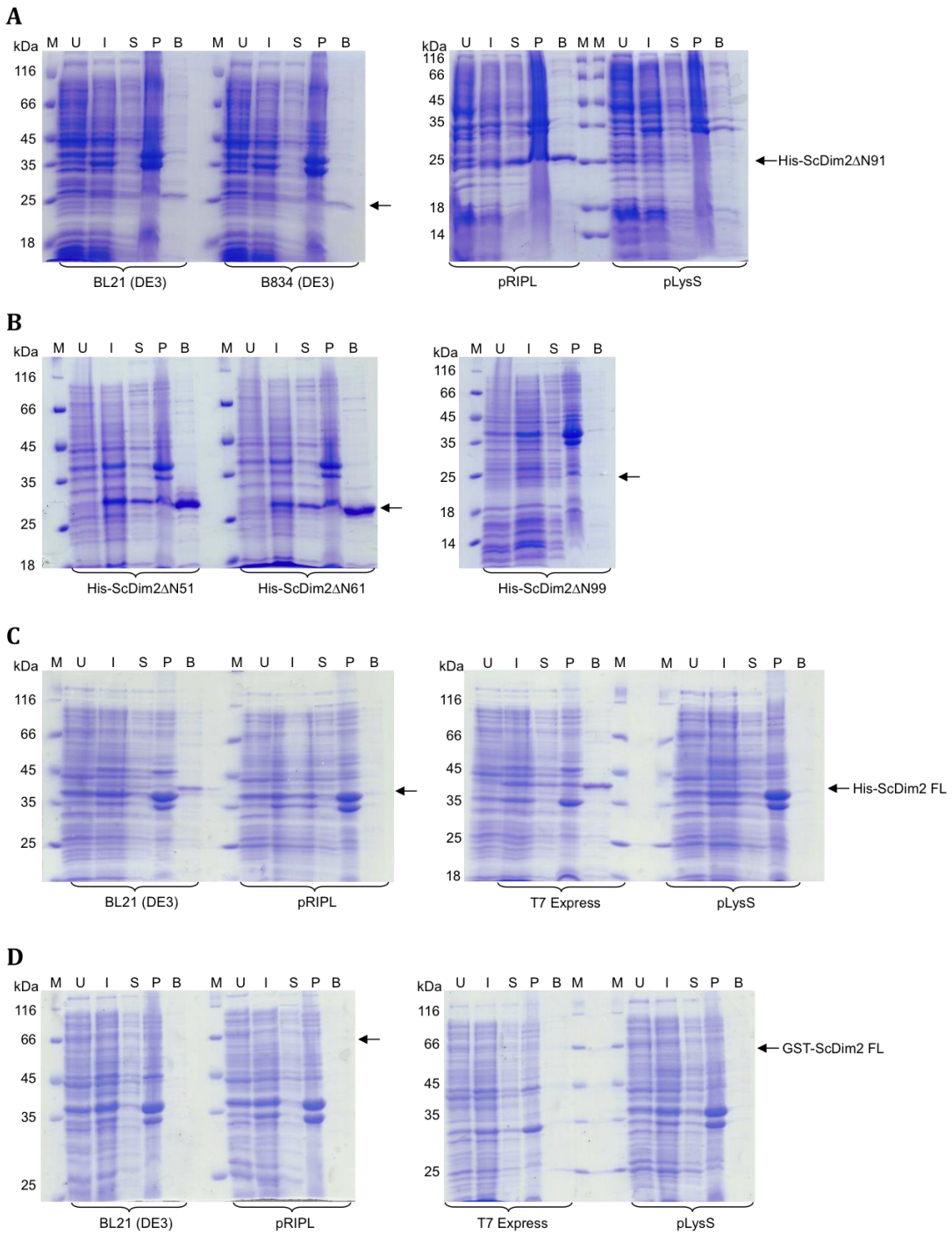


Figure 5.5: Expression trials for the different ScDim2 constructs created for this project.
The lane labelling is the same as for the previous figures.

A) Expression of the His-tagged ΔN91 construct. The protein expresses and is soluble in BL21 (DE3) and BL21 (DE3) pRIPL strains. A small amount was also observed using the B834 (DE3) cells.

B) Expression tests using the three new ΔN mutants (ΔN51, 61, and 99). The first two show good soluble expression in BL21 (DE3) cells in 2xTY medium. The longest truncation (ΔN99), on the other hand, shows no soluble expression at all.

C) Expression of the His-tagged full-length protein from the pEC-SHT plasmid. A small amount of expression is seen in BL21 (DE3) in 2xTY medium. The best expression is obtained using the T7 Express cell line. However, expression from the pEC-KHC plasmid was better.

D) The GST-tagged protein cloned into pEC-SGT is not expressed at all. Thus, it was not used.

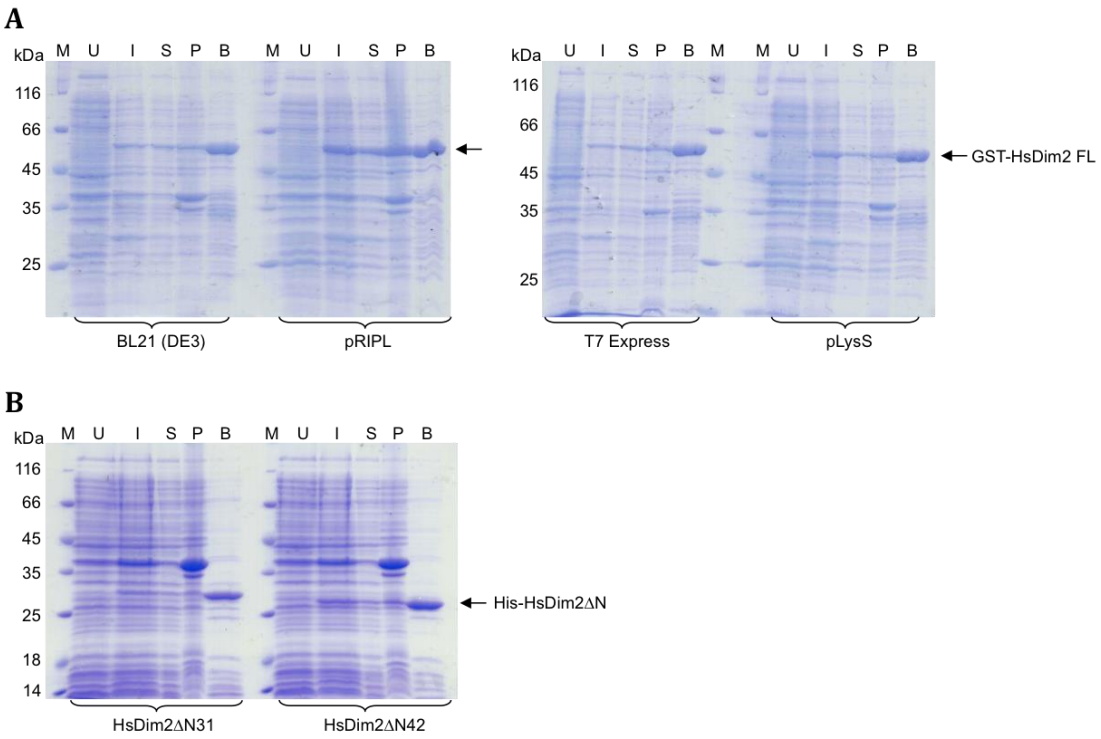


Figure 5.6: Expression trials of the human Dim2 (HsDim2) constructs.
The lane labelling is the same as for the previous figures.

A) The full-length GST-tagged protein was obtained from Dr. Nick Watkins. The HsDim2 over-expressed and was soluble in all cell strains used. They all also showed enrichment on the beads. The BL21 (DE3) strain was chosen for further cultures.

B) Expression of the two ΔN mutants created (ΔN31 and 42). They both showed soluble expression in BL21 (DE3) cells in 2xTY medium, which was enriched in the bead-bound fractions.

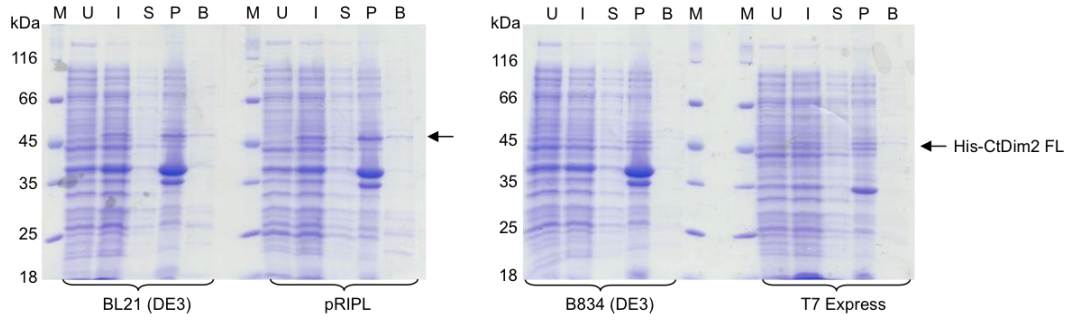


Figure 5.7: Expression trials of the codon optimised CtDim2 in 2xTY medium.
The lane labelling is the same as for the previous figures. A small amount of induced expression is seen in all strains apart from the B834 (DE3) strain. However, all of the protein is seen in the pelleted fraction showing that the protein is not soluble. Thus, this construct was not used any further.

5.4 Protein Purification

5.4.1 Initial trials for ScDim2 – full-length and ΔN91

Previous attempts to purify the full-length ScDim2 (performed by Iva Tchasovnikarova) were hampered by co-purification of the protein with ribosomal proteins and RNA derived from the expression host. Thus, the truncated version of the protein (ΔN91) was subjected to purification instead. However, after the first affinity step (**Figure 5.8A**), the protein precipitated upon GST tag cleavage and could not be resuspended with addition of salt. Thus, the truncated version only seemed to be soluble due to the GST fusion.

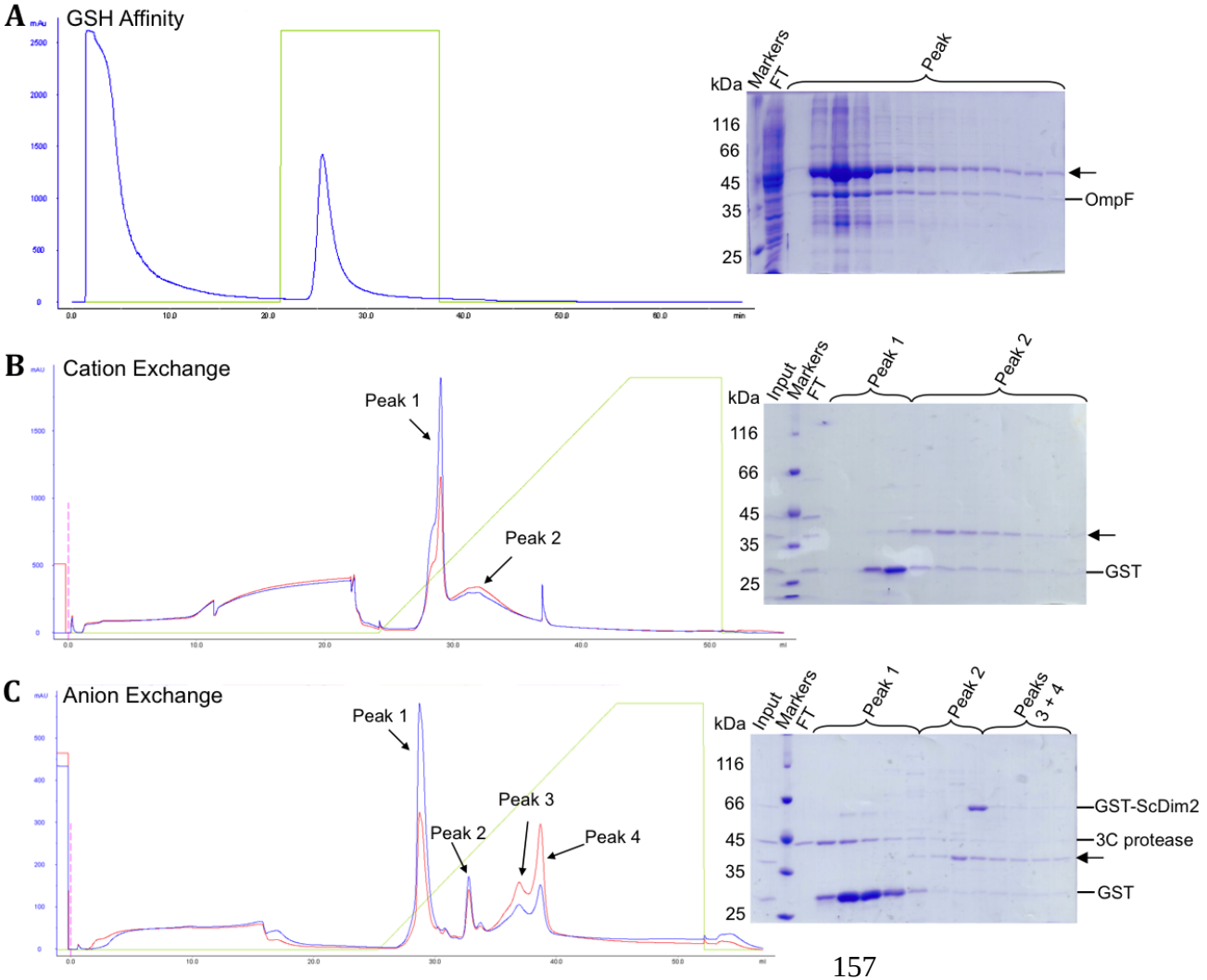
Going back to the GST-tagged full-length protein, alternative protocols that replaced the heparin sepharose step with ion exchange chromatography were tested. Neither Mono S nor Mono Q ion exchange columns were able to separate the protein from its contaminating RNA (**Figure 5.8B** and **C**).

To test whether Dim2 could be more easily purified in a complex with Nob1, a co-purification strategy was tested. GST-ScDim2 and His-ScNob1 were expressed separately and co-lysed. However, both proteins precipitated out of solution upon lysis and further purification could not be performed.

Finally, the His-ScDim2 construct was subjected to nickel affinity purification (**Figure 5.9**). Afterward, the RNA contamination was digested by incubation with Cyanase nuclease. This caused the protein to precipitate, suggesting that Dim2 is not stable in the absence of its RNA partner. Alternatively, the protein could protect a small portion of the RNA (the binding site) and precipitate due to the presence of a shorter nucleic acid.

Figure 5.8: Purification trials for ScDim2.

Absorbance profiles at 280 nm and 260 nm are shown in blue and red (where available) respectively. The protein of interest on the gel is indicated with an arrow.



A) The GST-tagged ScDim2 ΔN91 was affinity-purified using a column packed with glutathione resin. The protein was then eluted using a buffer containing 20 mM glutathione. The corresponding peak was analysed on SDS-PAGE (right image). The ScDim2ΔN91 protein co-eluted with *E. coli* OmpF protein (identified with MALDI-ToF mass spectrometry analysis). The truncated protein precipitated upon GST-tag cleavage and purification was discontinued.

B) GST-tagged full-length ScDim2 was purified as above giving the same elution profile. The GST-tag was cleaved using 3C protease. A Mono S cation exchange column was then loaded (result shown here) with the protein and eluted with a salt gradient (50 to 1000 mM). Some ScDim2 remained in the FT and the rest co-eluted with the cleaved GST. The FT was loaded onto a Mono Q column (anion exchange – see below).

C) Mono Q elution profile for ScDim2 using the same gradient as for Mono S. The column gives good separation but absorbance at 260 nm shows a lot of RNA contamination (peaks 3 and 4). A small portion appears relatively clean (peak 2) but it contains a large proportion of uncleaved GST-ScDim2.

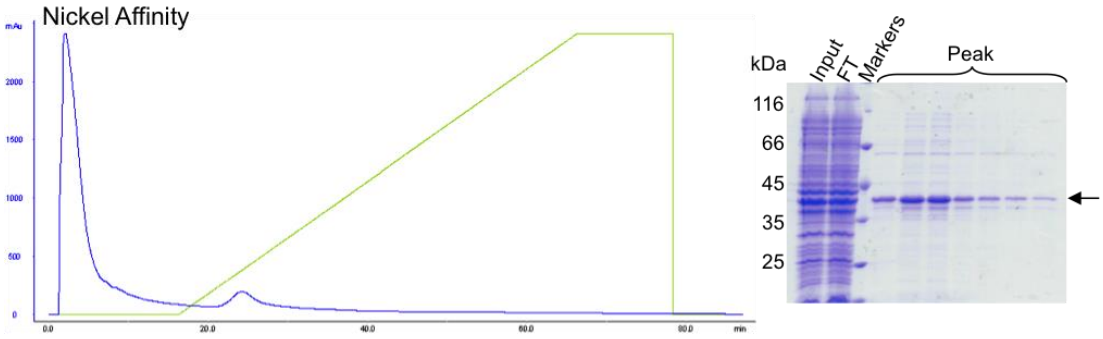


Figure 5.9: Nickel affinity purification of His-tagged ScDim2.
All the labelling and colour schemes are the same as on previous purification figures. The elution gradient ranged from 20 to 500 mM imidazole. The majority of the protein came in the flowthrough (FT) fraction as shown by the high absorbance before the start of the gradient and the SDS-PAGE gel on the right. The intense band (marked with an arrow) present in the input is still present in the FT. Western blotting (not shown) confirmed that the His tag is present but it appears to be inaccessible. Some protein still bound to the column and was eluted in the small peak at the start of the gradient. Fractions enveloping the peak were pooled and subjected to incubations with Cyanase nuclease.

5.4.2 HsDim2 – full-length

As previously mentioned, the HsDim2 construct was obtained from Dr. Nick Watkins. The Watkins lab had also established a single-step purification (GSH resin affinity for the GST tag), which included a detergent in the buffer. Two further steps were added to the purification in order to obtain pure HsDim2. After elution from the GSH resin, the GST tag was cleaved with 3C protease and the protein was dialysed into a low salt buffer. Then, the HsDim2 was injected onto a cation exchange column (pI of 9.8) and eluted with a salt gradient. Finally, the eluted peak was concentrated and injected onto a SEC column.

Figure 5.10 shows that the three purification steps worked well to give 2 mg of pure HsDim2 from 4 litres of culture. The protein eluted as a single peak in SEC showing no RNA contamination. The detergent appeared to be important in the first step as when it was excluded from the lysis buffer, the HsDim2 precipitated out of solution upon dialysis.

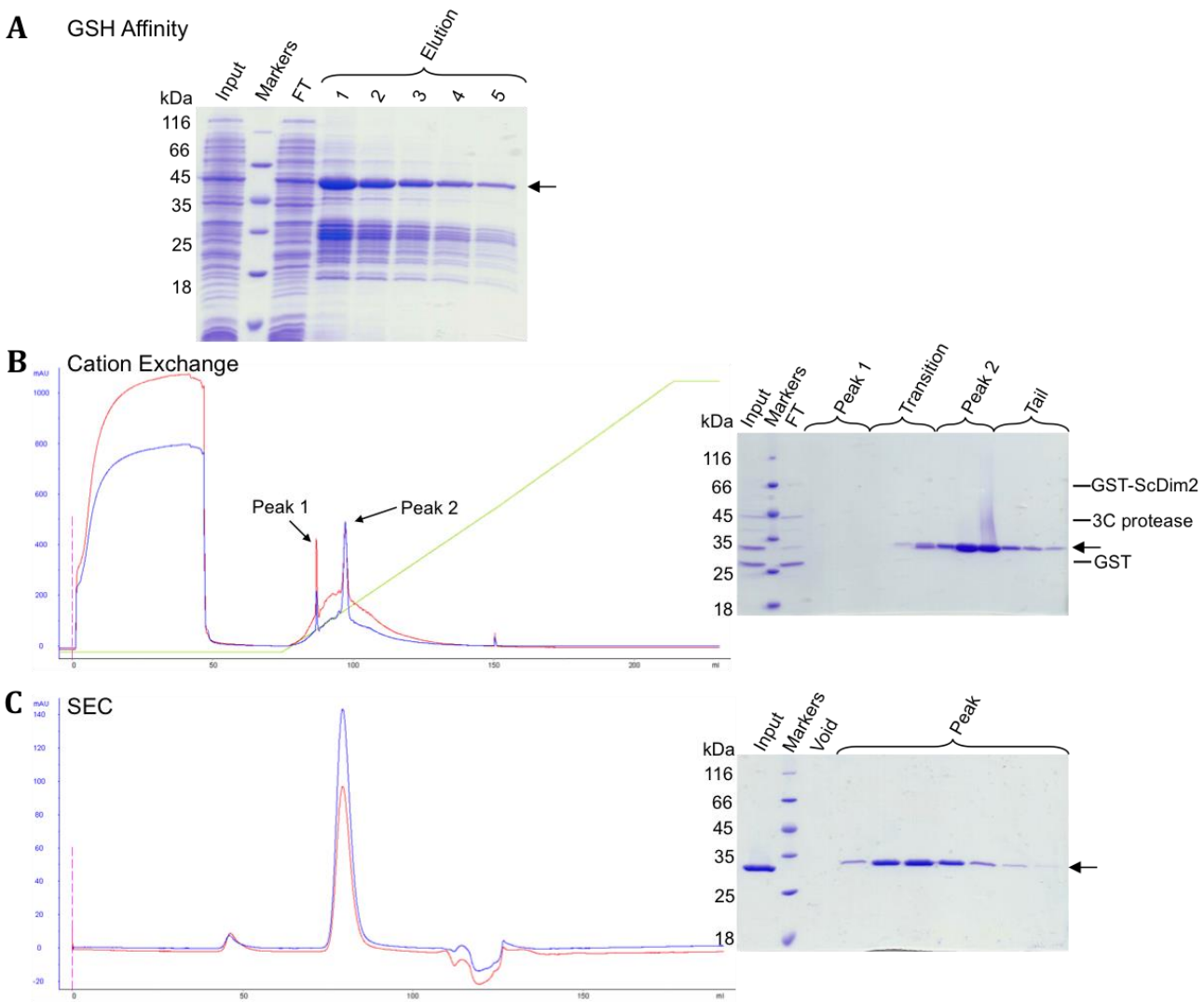


Figure 5.10: Purification of the GST-tagged HsDim2.
All labelling and colours are the same as on previous figures.
A) SDS-PAGE analysis of the elution from GSH resin. The protein was incubated with the beads and eluted with 20 mM glutathione in-batch. Five washes with the buffer were performed and all showed elution of the correct protein. The first four were pooled, dialysed, and used in the following steps.
B) The second step involved binding the protein to a Mono S cation exchange column and eluting it with a salt gradient (100 to 1000 mM KCl). The second peak was seen to contain the HsDim2 protein while the majority of RNA contamination eluted in the first peak. Thus, the second peak was pooled and concentrated in preparation for SEC.
C) The SEC profile for HsDim2. A single peak containing pure protein is seen. A small peak in the void is also seen. However, no bands are visible on the gel suggesting that it contains nucleic acids or other contaminants. Alternatively, the amount of protein in this peak is too small to be seen.

5.4.3 ScDim2ΔN51

Since the purification of HsDim2 worked well, the same protocol was used to purify the ScDim2ΔN51 truncated protein from yeast. However, the first step was replaced with nickel affinity as this construct contained a hexahistidine tag. As shown in **Figure 5.11**, 11 mg of pure protein were obtained from 2 litres of culture in the same manner as the human construct. However, it was found that the protein retained its ability to interact with nickel resin and it did not decrease in size observed by SDS-PAGE after incubation with the 3C protease. This suggests that the His tag could not be cleaved even though a cleavage site is present. The tag is likely to be too close to the folded regions of the protein and the cleavage site is obscured.

As the new protocol containing the detergent worked well for the human and yeast constructs, an attempt was made to go back to the full-length GST-ScDim2 and try this alternative purification method. However, when the protein was subjected to the second step (cation exchange), it was found to co-elute with the GST tag and 3C protease, which normally do not bind to this column (see HsDim2 purification). Also, significant RNA contamination was seen (**Figure 5.12**).

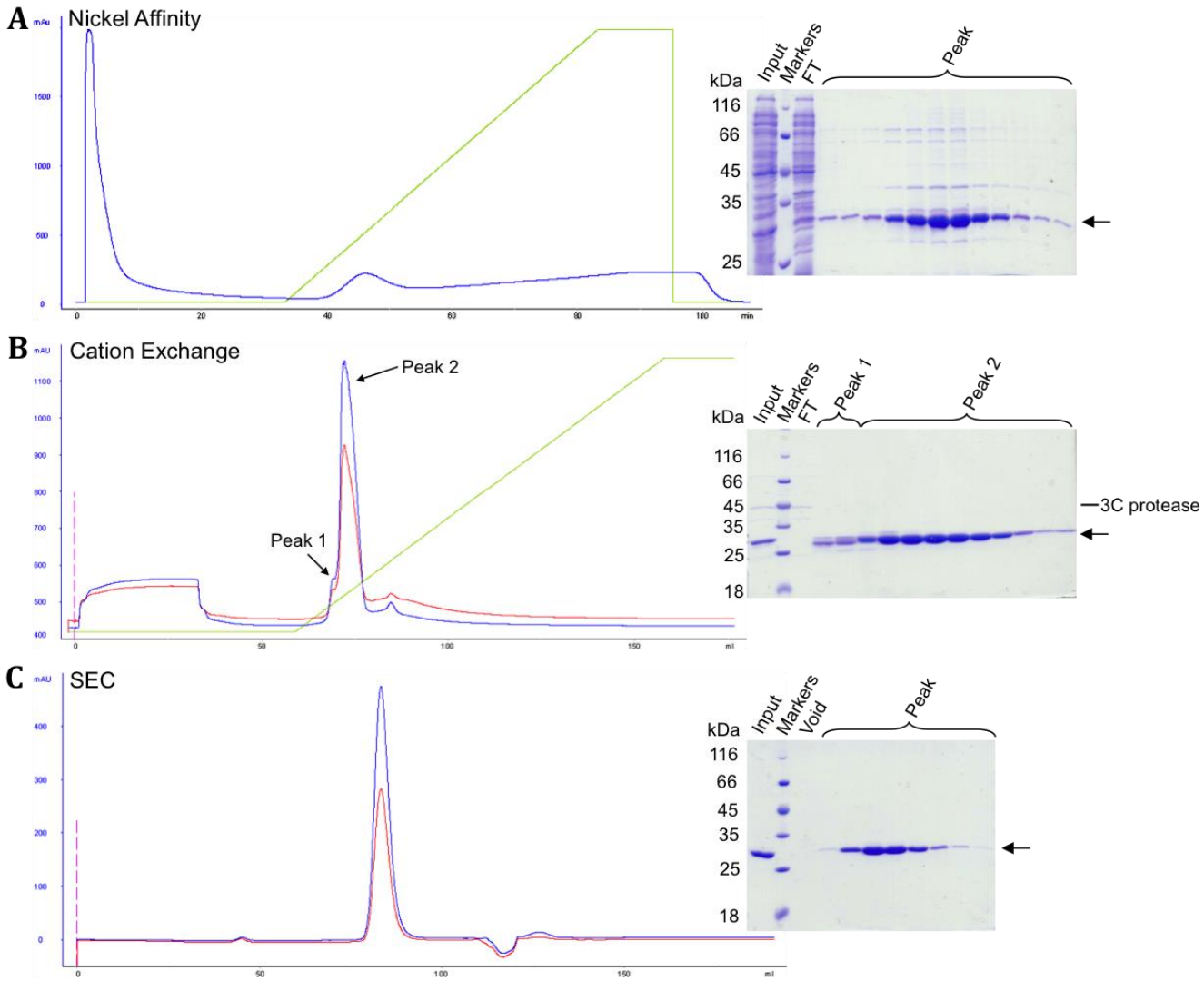


Figure 5.11: Purification of the His-tagged ScDim2ΔN51 protein.
The colours and labels are the same as in previous figures.
A) Nickel affinity purification. The protein was eluted with an imidazole gradient (20 to 500 mM) and was seen as a single peak. The fractions containing this peak were pooled and dialysed overnight.
B) Mono S cation exchange purification was then performed. Pure protein was seen to elute in the second peak. An overlapping peak containing some of the contaminants was also present. The second peak was pooled and concentrated for SEC.
C) The SEC elution profile for ScDim2ΔN51. A very small peak in the void (i.e. FT) is seen. However, all of the protein elutes as a single peak. The fractions containing this peak were pooled and concentrated.

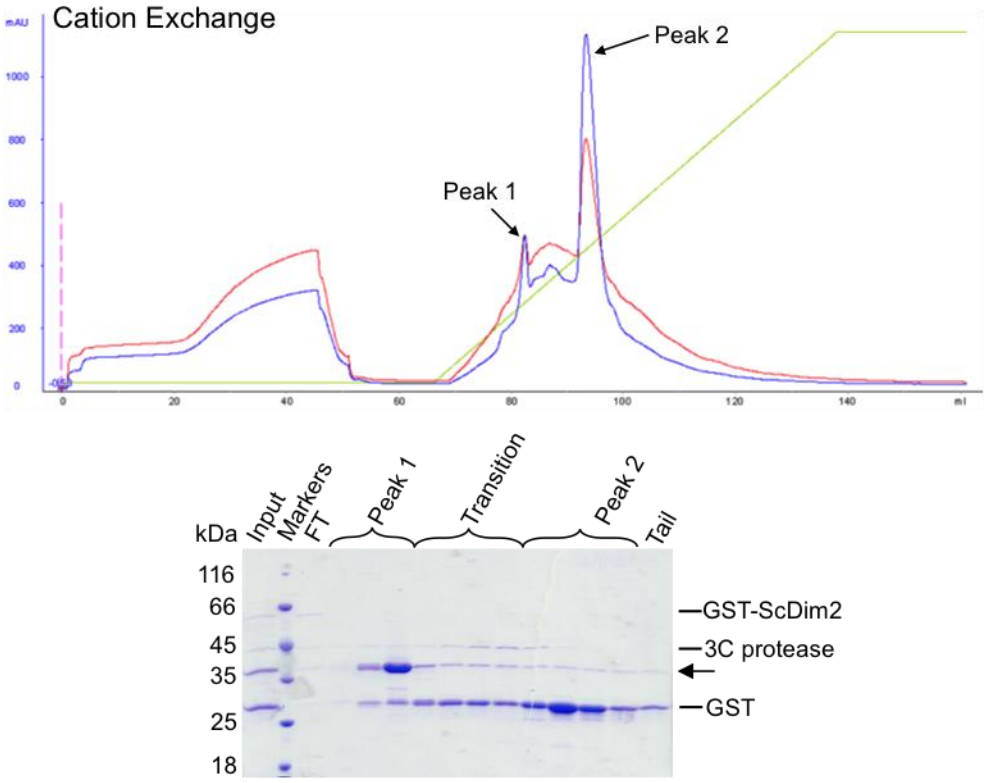


Figure 5.12: Purification of full-length GST-tagged ScDim2 using the protocol established for the human protein.

The labels and colours are the same as on previous figures. The first step using in-batch affinity purification with GSH resin looked the same as for HsDim2. The protein was pooled and dialysed overnight. Then, it was injected onto a Mono S cation exchange column (chromatogram shown here). The results above show that ScDim2 co-elutes with GST and 3C protease, which normally do not bind to this column (as seen with HsDim2). This suggests that the protein forms a large complex/precipitate with GST and/or 3C protease. Also, the absorbance at 260 nm is relatively high compared to 280 nm in peak 1 (where the ScDim2 elutes) suggesting RNA contamination.

5.5 Oligomerisation State of Dim2 Remains in Question

Two constructs of Dim2 were successfully purified: the full-length human protein and the N-terminally truncated yeast protein. Their elution from the SEC column suggests that they both might form dimers (**Figure 5.13**). However, as stated before, SEC profiles depend on the shape as well as size of the protein. Thus, it is possible that it forms an elongated monomer instead of a dimer. In order to determine which hypothesis is more likely, the hydrodynamic radii (R_h) and radii of gyration (R_g) were calculated for existing solved structures using the HYDROPRO software (Ortega et al, 2011).

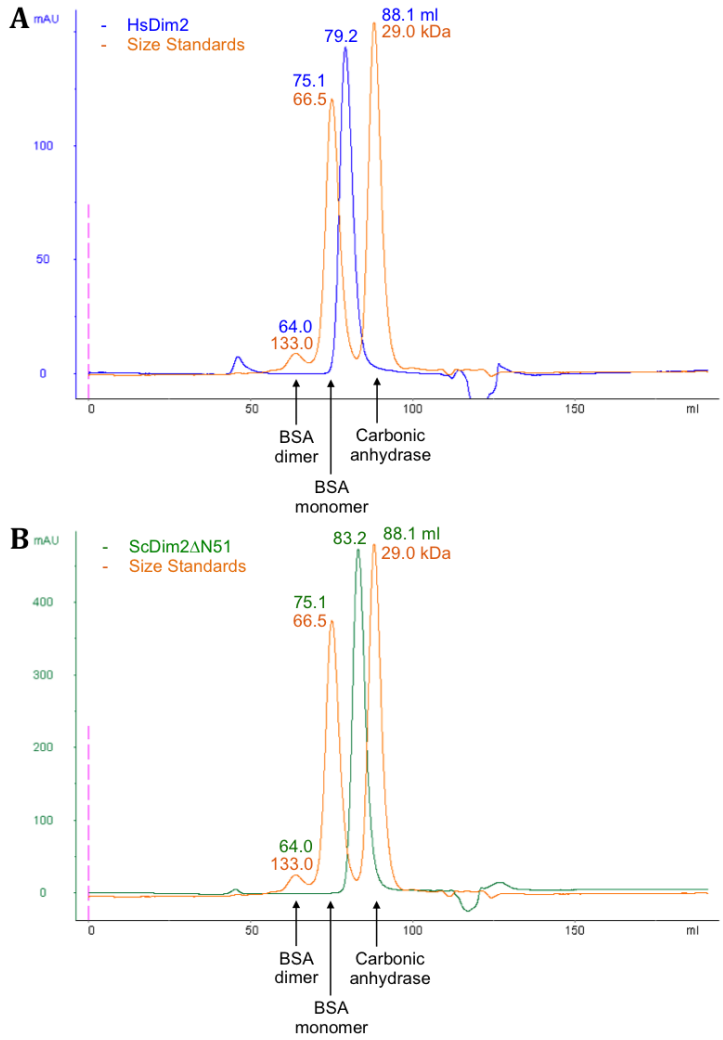


Figure 5.13: HsDim2 and ScDim2ΔN51 SEC results compared to size standards. The protein standards are labelled under the chromatogram. Their sizes (in kDa) are shown in orange next to each peak with the corresponding elution volumes shown above these in blue or green. **A)** Comparison of HsDim2 with the size standards. HsDim2 appears smaller than the BSA monomer but larger than the carbonic anhydrase. However, the predicted size is only 28.3 kDa. **B)** SEC results for ScDim2ΔN51 compared to the standards. This protein is slightly smaller than HsDim2 but still elutes before the carbonic anhydrase. The predicted size is only 27.4 kDa.

For this purpose, the PDB files for the size standards used in SEC and the archaeal *P. horikoshii* Dim2 were used as described in the methods. The results of the analysis are summarised in **Table 5.2**. The hydrodynamic radius is a measure of the size of the protein. However, it assumes a spherical shape. The radius of gyration is more dependent on the structure as it reflects the root mean square distance from the centre of the molecule to each element. The ratio of the two values (R_g/R_h) gives an assessment of the globularity of the species. For a spherical protein, this ratio should be 0.775 (Burchard et al, 1980; Vandesande & Persoons, 1985). A higher ratio corresponds to an elongated protein (R_g is larger than for a spherical protein) whereas a lower one shows that the density of the molecule is skewed towards the core (R_g is smaller than for a globular species).

Table 5.2: Summary of results obtained from HYDROPRO for the different structures and the elution volumes obtained from SEC.

Protein	PDB ID	MW (kDa)	R_h (nm)	R_g (nm)	R_g/R_h	Elution Volume (ml)
BSA monomer	4F5S	66.4	3.60	2.75	0.765	75.0
Carbonic Anhydrase	1V9E	29.0	2.48	1.84	0.744	88.1
Archaeal Dim2	2E3U	25.1	2.32	1.82	0.785	-
ScDim2ΔN51	-	27.4	-	-	-	82.1
HsDim2	-	28.3	-	-	-	79.5

The results obtained from HYDROPRO show that while the two size standards (BSA and carbonic anhydrase) are dense in the core with a couple of extremities (R_g/R_h ratio below 0.775), the solved archaeal Dim2 structure is elongated (ratio of 0.785). If the eukaryotic structure is similar to the archaeal one, Dim2 should elute marginally earlier in SEC than would be predicted from its molecular weight. However, the ratio is still relatively close to that for a globular protein and the shape alone cannot explain the large discrepancies in elution volumes. For example, HsDim2 elutes closer to BSA than carbonic anhydrase even though it is smaller in size than the latter protein. Thus, it is more likely that the proteins form a more globular dimer than a largely elongated monomer.

In the future, SEC-MALS should be performed to determine the true oligomerisation state. Then, DLS could be used to experimentally obtain the hydrodynamic radii and information about the diffusion of the eukaryotic recombinant proteins in solution.

5.6 Determination of Protein Stability by Thermal Denaturation

5.6.1 HsDim2

In order to determine if the purified HsDim2 protein is stable, thermal denaturation was performed. Initially, a trial using different concentrations (1-10 µM) of HsDim2 was performed. This experiment showed that, unlike Nob1, HsDim2 is seen to be more stable at higher protein concentrations. The melting temperature dropped from 48.7˚C at 10 µM to 43.5˚C at 1 µM. A buffer trial was performed using 3 µM HsDim2 (**Figure 5.14**), as this was the lowest concentration predicted to give a good signal-to-noise ratio.

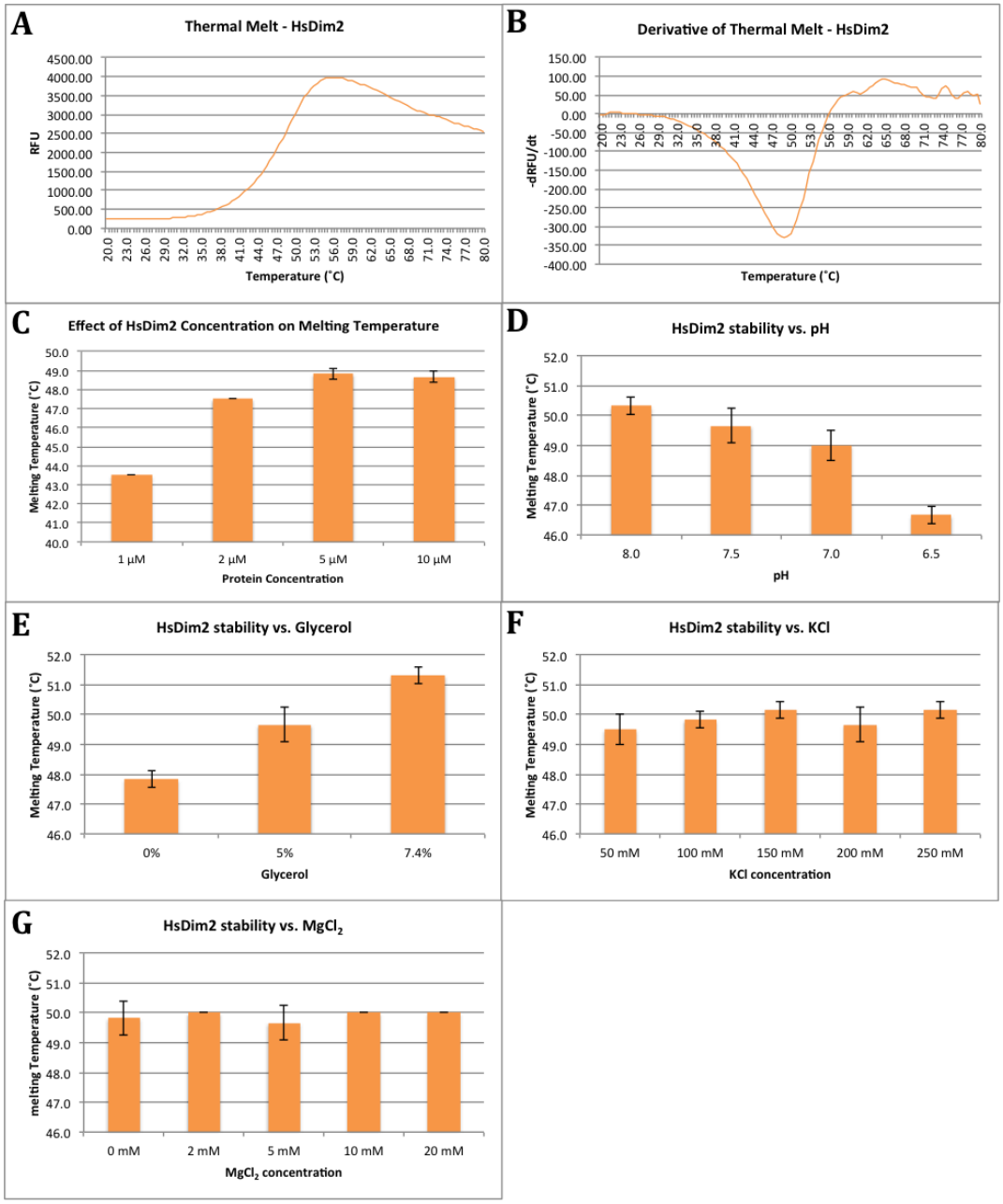


Figure 5.14: Thermal denaturation results for HsDim2.

The experiments were performed in triplicate. The error bars represent the standard deviations.

A) An example of the raw thermal melt obtained. The increase in relative fluorescence units (RFU) was recorded over changes in temperature. A single melt curve is seen.

B) An example derivative of the thermal melt. The bottom of the peak corresponds to the melting temperature (T_m). This value was taken for each of the experiments and averaged over the three repeats to obtain the summarised results.

C) Summary of the effect of protein concentration on protein stability. Higher HsDim2 concentrations led to higher melting temperatures suggesting increased stability.

D) 3 µM protein was used for all the following experiments. Summary of the effect of decreasing pH on the melting temperature. The protein appears to be more stable at higher pH.

E) Effects of glycerol concentration on HsDim2 stability. The T_m increases with glycerol.

F) KCl concentration has no significant effect on protein stability.

G) $MgCl_2$ has no significant effect on HsDim2 stability.

It was found that the pH of the buffer had a significant effect on the melting temperature of HsDim2 (50.3˚C at pH 8.0 to 46.7˚C at pH 6.5). Thus, the protein was only stored in buffers at pH 8.0 or 7.5 in later experiments. As with Nob1, it was also found that the stability increases with glycerol concentration. Salt and magnesium concentrations had no significant effect on the melting temperature suggesting that they do not affect protein stability.

Due to time constraints, the thermal denaturation assay was not performed for ScDim2ΔN51. However, it should be tested in the future.

5.7 Identification of Stable Domains by Limited Proteolysis

5.7.1 HsDim2

Since the full-length human Dim2 purified well and was predicted to have two domains (the KH and KH-like domains), limited proteolysis was performed to see which parts were stably folded. The results, shown in **Figure 5.15**, suggest that the protein is relatively stable as only the highest concentrations of proteases show full digestion of the protein. A single band in the GluC 1:10 lane was observed. A band of the same size was also seen using 1:10 Elastase (and partially in 1:100 Trypsin). Thus, these two enzymes were chosen for a time trial (at a 1:100 dilution).

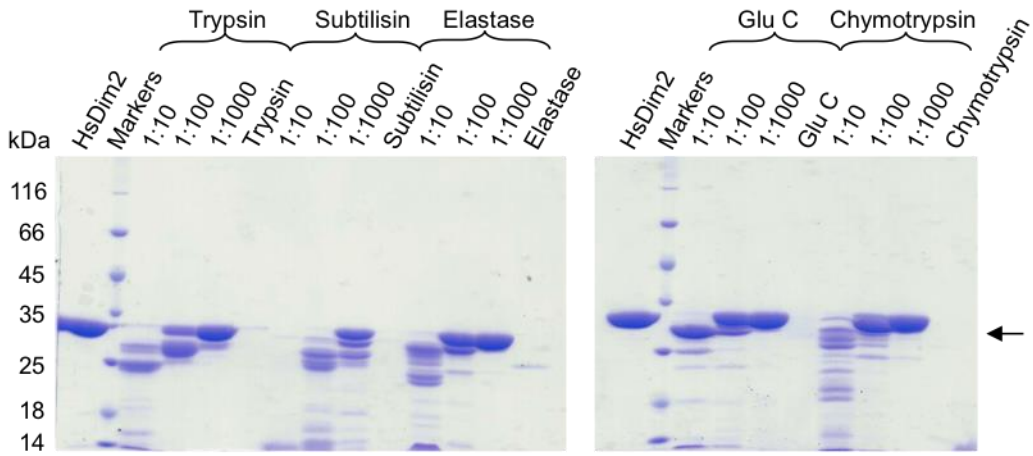


Figure 5.15: Limited proteolysis of HsDim2.
The enzymes used are shown above. Three dilutions per enzyme (of 1 mg/ml stocks) were used. Limited proteolysis allowed mapping of stable regions. The band of interest in the 1:10 GluC lane is marked with an arrow. The experiment was later scaled up in order to obtain a large amount of the truncated protein.

Figure 5.16A shows that, after 4 hours, some full-length protein is still present. Thus, an overnight incubation was also tested. The long incubation with GluC produced, almost exclusively, the band of interest (**Figure 5.16B**). This enzyme and incubation time were chosen for preparative proteolysis. MALDI-ToF mass spectrometry analysis of the band identified it as an N-terminal truncation (**Figure 5.17**).

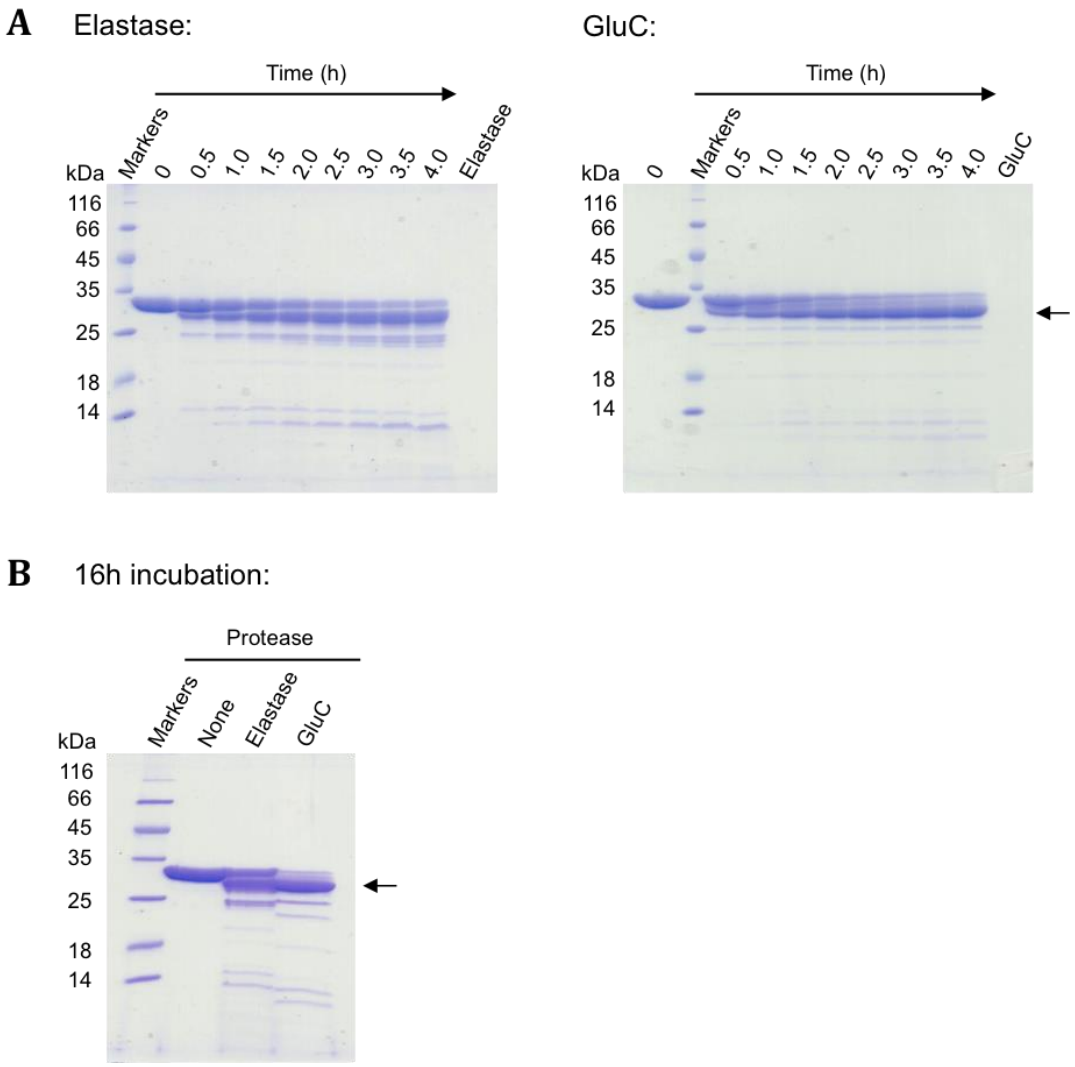


Figure 5.16: Proteolysis of the HsDim2 protein over time using 1:100 Elastase or GluC.

A) Time trials using 1:100 Elastase (left) or GluC (left). The protein was incubated with each protease for up to 4 hours. Both proteases showed the production of the band of interest. However, the full-length protein was still visible. Thus, overnight incubation was also performed.

B) Results of the overnight proteolysis. The Elastase lane still shows a significant amount of the full-length protein. On the other hand, GluC proteolysed almost all of the protein. Thus, this enzyme was chosen for large-scale preparative proteolysis.

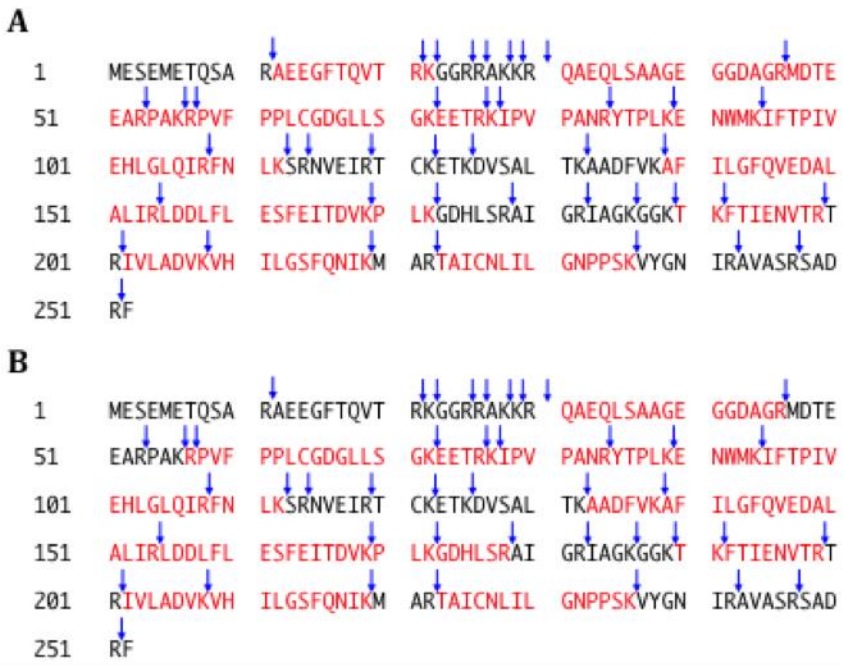


Figure 5.17: MALDI-ToF analysis results for the full-length and proteolysed HsDim2.
The band was excised from the gel and subjected to trypsin digestion. The predicted proteolytic sites are indicated with the blue arrows. The peptides identified by mass spectroscopy are shown in red.

A) Results for the full-length protein. Peptides covering almost all of the protein were identified. The ones not seen appear to be too short.

B) Results for the proteolysed fragment of HsDim2. All of the peptides except the N-terminal one have been identified. This suggests that the truncation occurs in the N-terminus of the protein.

Preparative proteolysis was performed on fresh HsDim2 purified using the first two steps described previously. After cation exchange, the protein was mixed with 0.005 mg GluC per 1 mg of HsDim2 and incubated at 4˚C overnight. The reaction was then stopped with protease inhibitors and injected onto a SEC column. Two overlapping peaks were observed in the elution profile, containing different truncation products (**Figure 5.18**). They were pooled separately and concentrated. The longer fragment gave a yield of about 400 µg while the shorter one only 300 µg from 4 litres of culture. They were then used for crystallisation screening. No hits were obtained using these fragments.

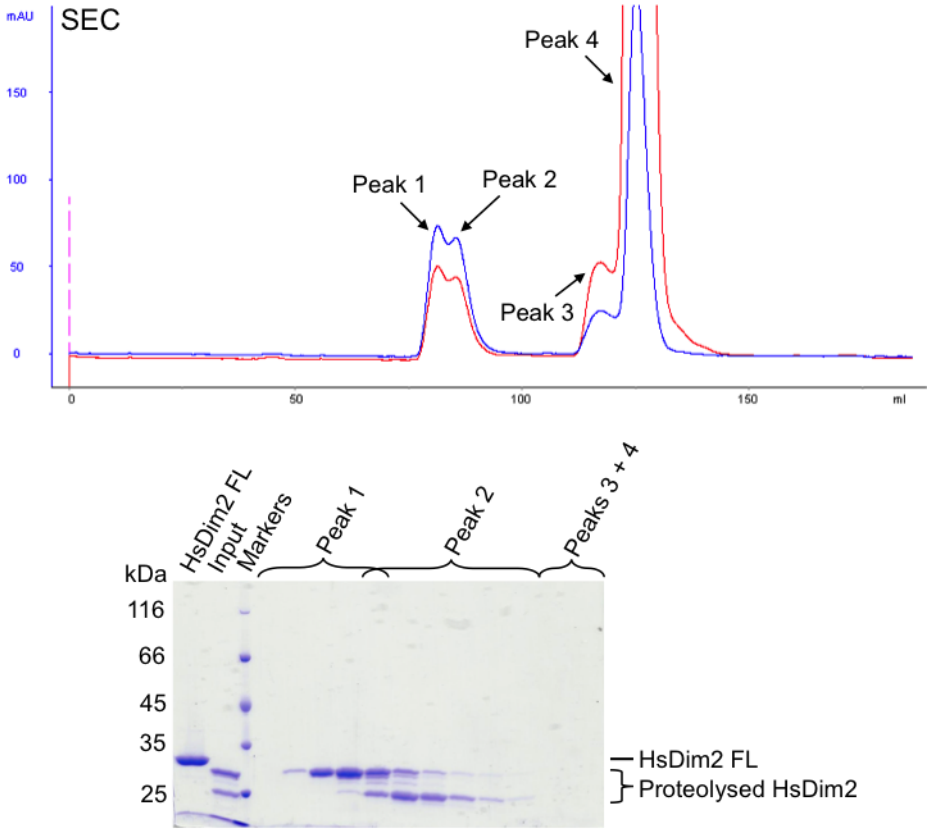


Figure 5.18: Preparative proteolysis and SEC purification of the HsDim2 protein.
Absorbance at 280 nm is shown in blue while 260 nm is shown in red. Two overlapping peaks were eluted from the SEC column. SDS-PAGE analysis of the fractions identified two truncation products. They were pooled separately and used for crystallisation screening. Peaks 3 and 4 likely contained the excess protease inhibitors. This is supported by the small size of the particles and the high absorbance at 260 nm. Also, SDS-PAGE shows that there are no proteins present in these peaks.

5.8 Testing Interactions with Different Binding Candidates

5.8.1 ScNob1 and ScDim2ΔN51 do interact

Multiple studies have found that Nob1 and Dim2 interact. To test whether this is true for recombinantly purified proteins, semi-analytical SEC was performed using the two yeast proteins. The elution profiles looked additive and the individual peaks were visible suggesting a lack of interaction (**Figure 5.19**). However, when the experiment was repeated at higher pH (**Figure 5.20**) or higher salt concentrations (**Figure 5.21**), a shift to an earlier elution volume was observed. Complex formation

was confirmed using SDS-PAGE of the fractions. Nevertheless, free proteins are still seen along with the complex.

In the future, other quantitative binding experiments should be performed (such as ITC) to obtain more information about the interaction, such as the dissociation constant and the molar ratio.

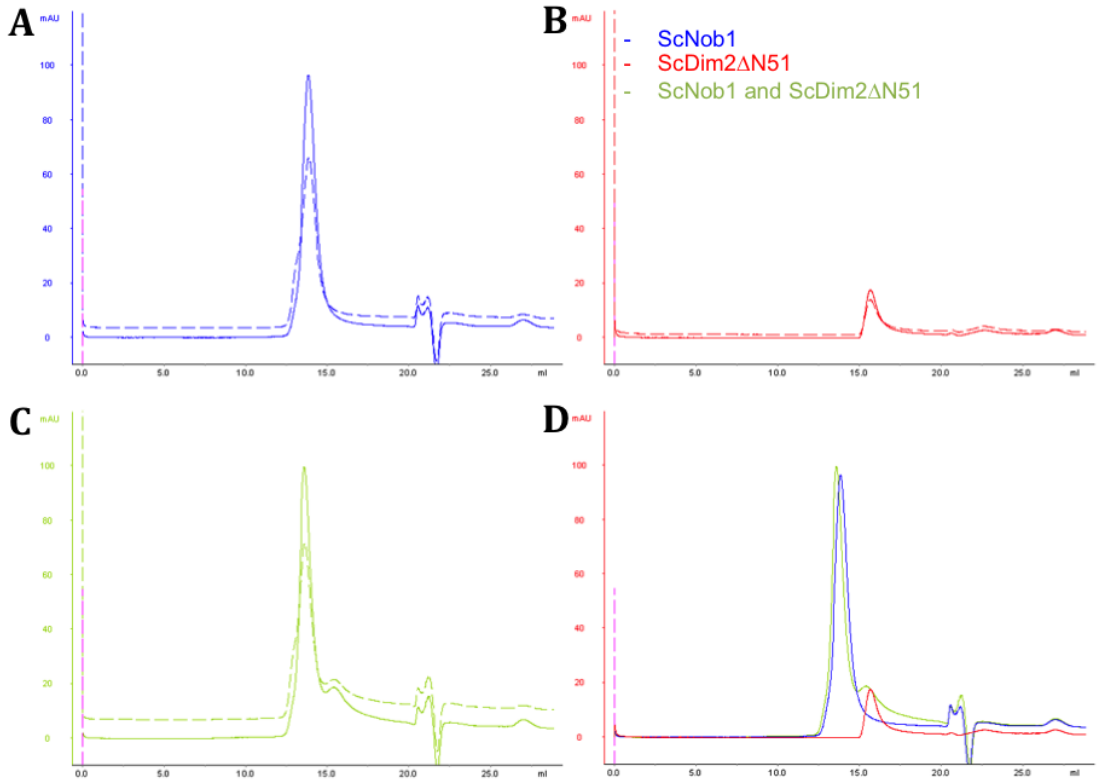


Figure 5.19: Semi-analytical SEC results using ScNob1 and ScDim2ΔN51.
Absorbance at 280 nm is shown as a solid line whereas absorbance at 260 nm is shown as a dashed line. The key for the colours is shown in the top right side.

A) The SEC profile for ScNob1 only. As before, a single peak is seen.

B) The chromatogram for ScDim2ΔN51 alone.

C) The SEC result when ScNob1 is mixed with ScDim2ΔN51 in a 1:1 ratio. The two peaks remain separate, although, as with HsDim2, they overlap making them difficult to define.

D) An overlay of the three previous results showing absorbance only at 280 nm. The top of the peak has shifted slightly to an earlier elution volume suggesting a weak interaction. However, this would correspond to a very small increase in size. Thus, it is possible that it is caused by experimental error.

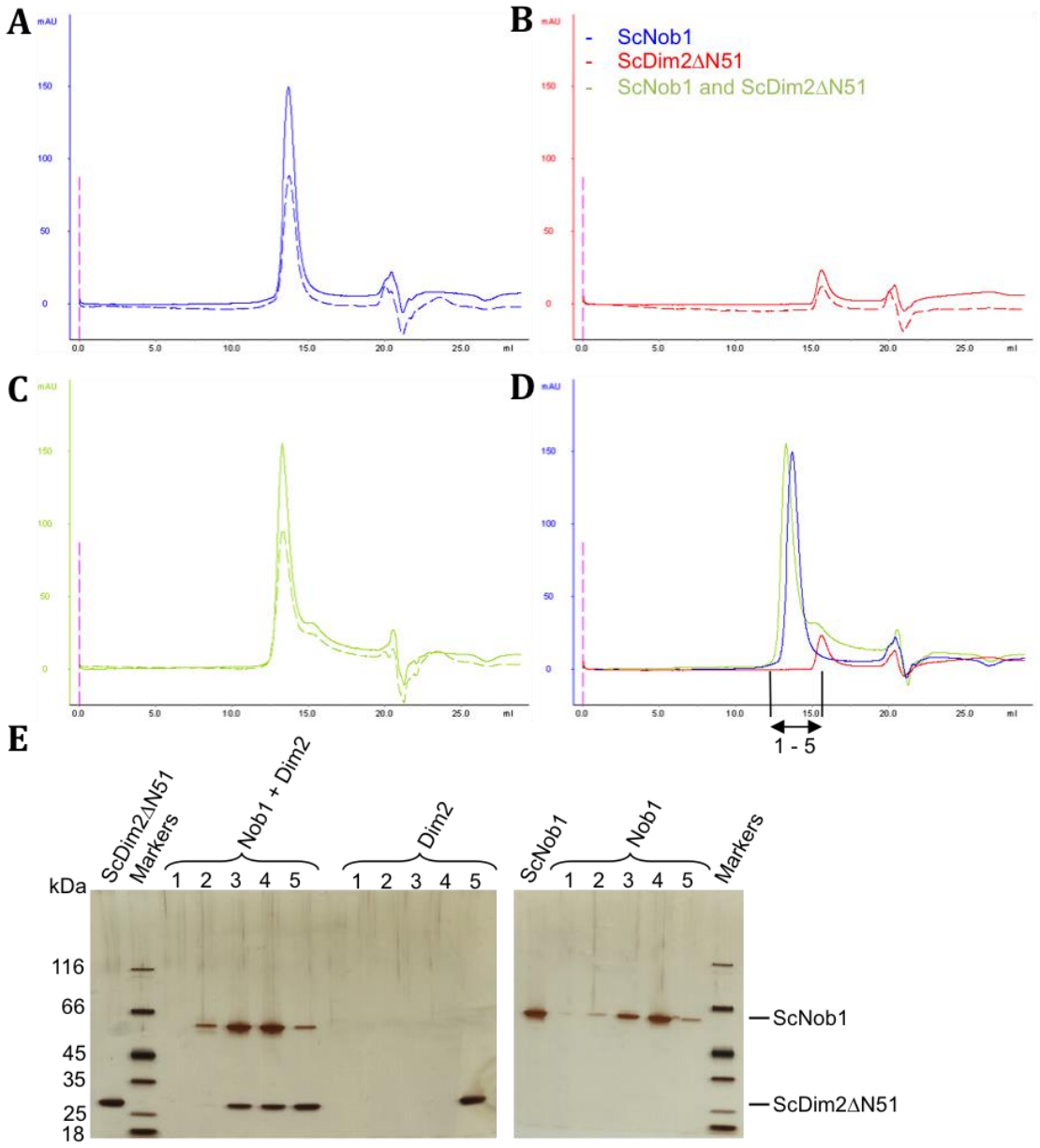


Figure 5.20: Semi-analytical SEC of ScNob1 and ScDim2ΔN51 at high pH (8.5 instead of 7.5). Absorbance at 280 nm is shown as a solid line whereas absorbance at 260 nm is shown as a dashed line. The key for the colours is shown on the right.

A) The SEC profile for ScNob1 only.

B) The chromatogram for ScDim2ΔN51 only. The peak looks the same as at pH 7.5.

C) The SEC result when ScNob1 is mixed with ScDim2ΔN51 in a 1:1 ratio. Separate profiles for the two peaks are still seen. However, the larger peak is visibly shifted to an earlier elution volume.

D) An overlay of the three previous results showing absorbance only at 280 nm. As mentioned above, the peak is definitely shifted to the left supporting complex formation. However, a peak for ScDim2ΔN51 alone is still visible.

E) Silver stained SDS-PAGE analysis of the fractions from each experiment indicated with an arrow in panel **D**. The distribution of the two proteins changes (to the left) when they are mixed together. This suggests that the two protein co-elute supporting that a complex is formed. However, not all of the free protein is involved.

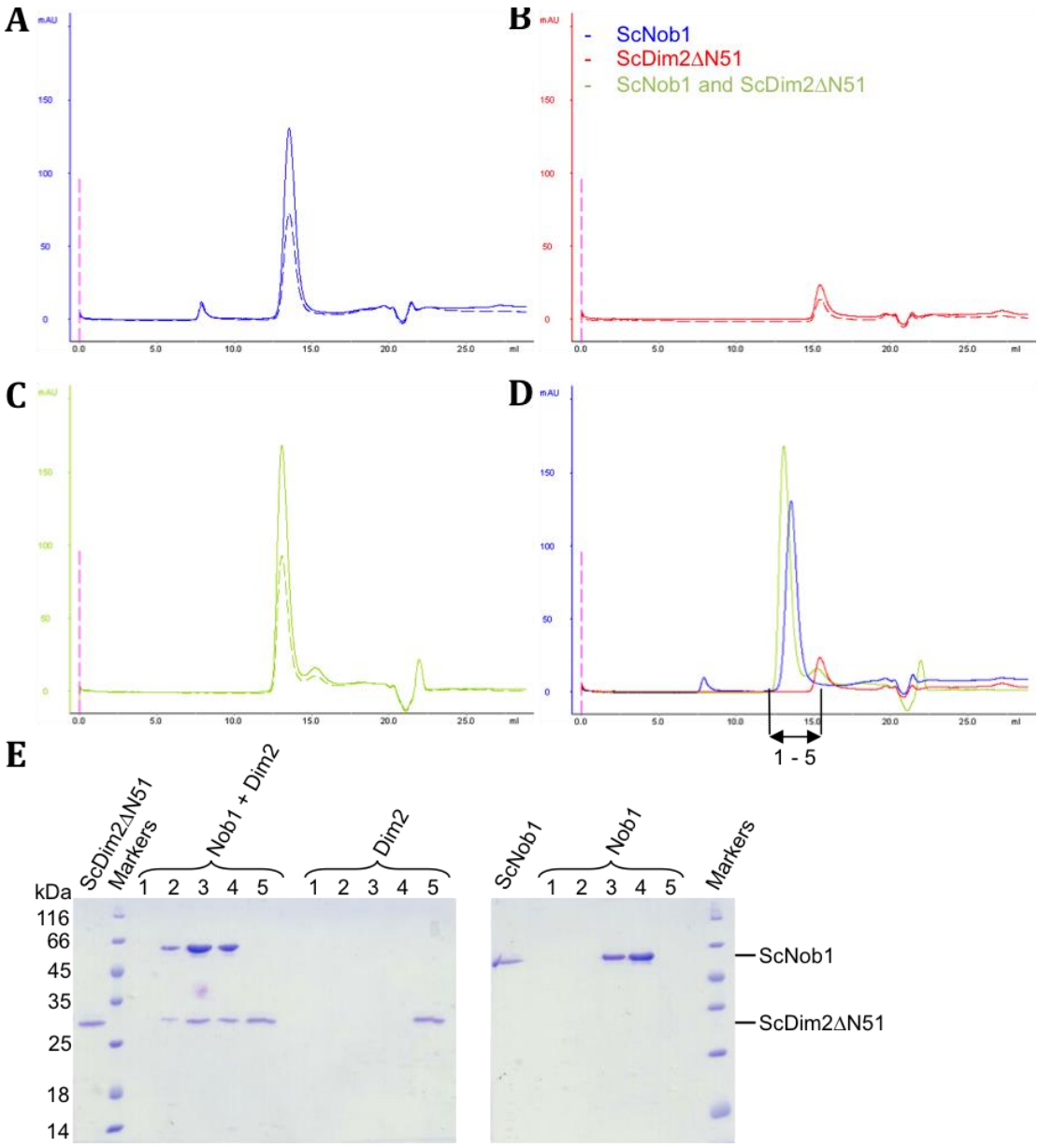


Figure 5.21: Semi-analytical SEC analysis of ScNob1 binding to ScDim2ΔN51 at higher salt concentrations (350 mM instead of 150 mM KCl).

Absorbance at 280 nm is shown as a solid line whereas absorbance at 260 nm is shown as a dashed line. The key for the colours is shown in the top right corner.

A) The SEC profile for ScNob1 only.

B) The chromatogram for ScDim2ΔN51 alone.

C) The SEC result when ScNob1 is mixed with ScDim2ΔN51 in a 1:1 ratio. The larger peak is seen to shift to an even earlier elution volume although a free ScDim2ΔN51 peak is still seen.

D) An overlay of the three previous results showing absorbance only at 280 nm. As with higher pH, higher salt is seen to cause the peak to shift to the left. More of the ScDim2ΔN51 peak also appears to be incorporated into the other peak.

E) Coomassie stained SDS-PAGE analysis of the fractions indicated in panel **D** from each experiment. Again, when the two proteins are mixed together, the distribution of protein shifts to the left. Thus, a complex is definitely formed.

5.8.2 ScNob1 and HsDim2 do not interact

The semi-analytical SEC was then repeated using yeast Nob1 and human Dim2. The Dim2 protein is highly conserved between yeast and human, particularly in the KH domain (see **Figure 1.7**). The surface forming interactions with Nob1 should also be highly conserved to preserve the mechanism of complex formation. However, as shown in **Figure 5.22**, no interaction was observed between the recombinant proteins. It is noteworthy that this experiment was only performed using the initial conditions. A higher pH or salt concentration could show an interaction between these two constructs.

5.8.3 ScDim2ΔN51 does not interact with ScDim1 *in vitro*

A study using co-precipitation from cell extracts has also found that Dim2 interacts with Dim1 in the pre-ribosome (Vanrobays et al, 2004). Again, this was tested with the recombinant proteins *in vitro*. ScDim1 was previously purified by Anees Basha. This was mixed in a 1:1 ratio with ScDim2ΔN51 and injected onto a semi-analytical SEC column. No interaction was observed in these conditions (**Figure 5.23**).

As a negative control, the same experiment was repeated with ScNob1 and ScDim1. There is no evidence that these two proteins directly interact during ribosome biogenesis. The result also looks additive (**Figure 5.24**), suggesting that neither of the proteins interacts with Dim1. However, the ScDim1 protein was kept in the freezer as a glycerol stock for a substantial length of time and could have deteriorated during storage. Also, the experimental conditions might not have been ideal for complex formation.

5.8.4 ScNob1 and ScFap7 do not interact in semi-analytical SEC

Recently, it has been shown that Nob1 can interact with Fap7 in archaea. Thus, ScFap7 was purified by Dr. Uma Jayachandran and semi-analytical SEC was used to see if these results could be replicated using yeast proteins. However, as shown in **Figure 5.25**, no interaction between the two proteins was seen in the conditions tested.

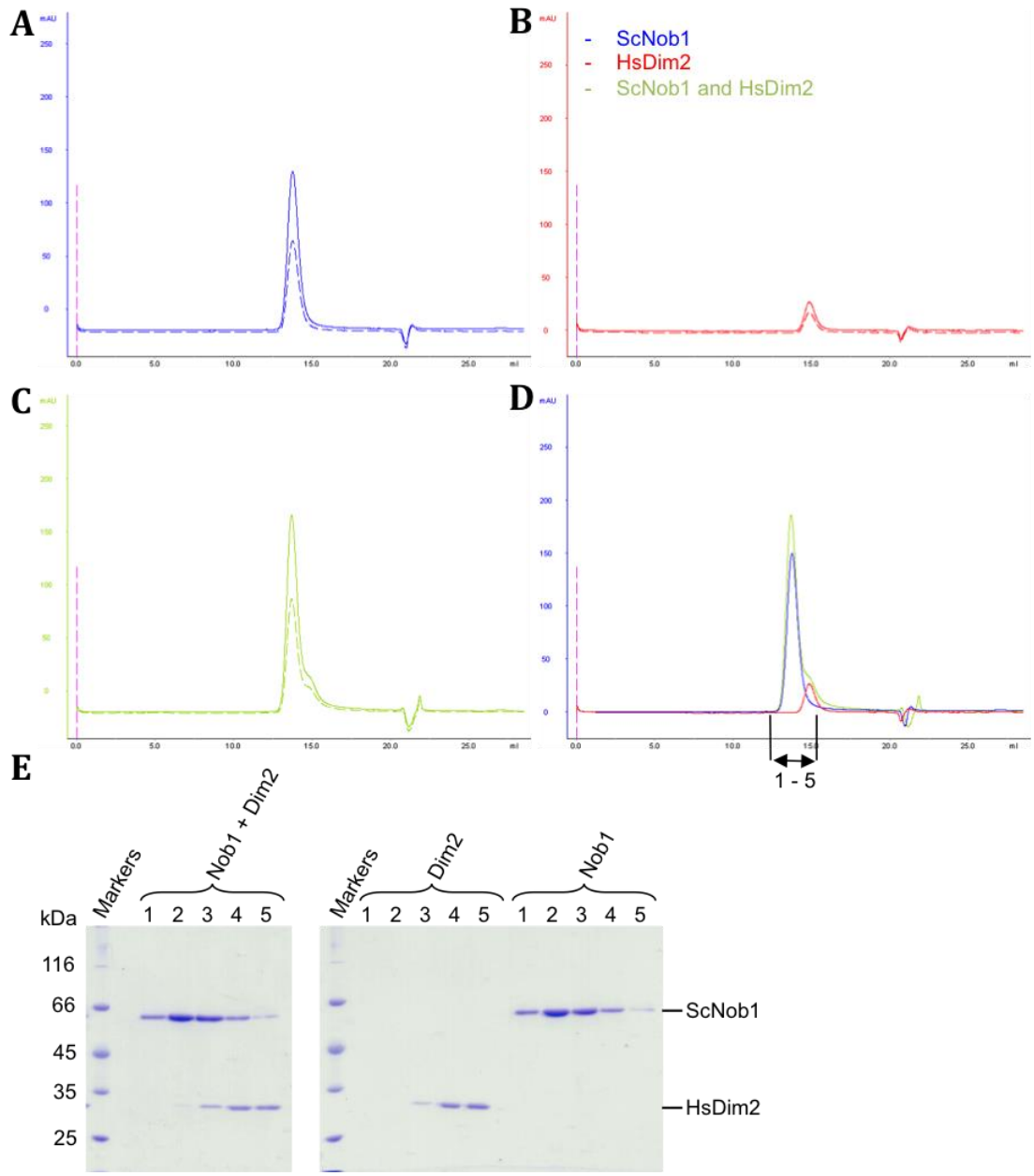


Figure 5.22: Semi-analytical SEC results for ScNob1 mixed with HsDim2.
Absorbance at 280 nm is shown as a solid line whereas absorbance at 260 nm is shown as a dashed line. The key for the colours is shown on the top right.

A) The SEC profile for ScNob1 only. A single peak is seen.

B) The chromatogram for HsDim2 alone. Again, a single peak is observed.

C) The SEC result when ScNob1 is mixed with HsDim2 in a 1:1 ratio. The two peaks appear to be separate.

D) An overlay of the three previous results showing absorbance only at 280 nm. The tops of the two peaks are in the same positions on the elution profile suggesting that there is no interaction. A slight increase in absorbance can be seen. This is likely due to pipetting or injection variation.

E) Coomassie stained SDS-PAGE analysis of the fractions containing the peaks (the fractions spanned the elution volumes indicated on panel **D** with the arrow). The two proteins alone and mixed together show the same distribution of protein in the peak fractions, supporting that there is no interaction.

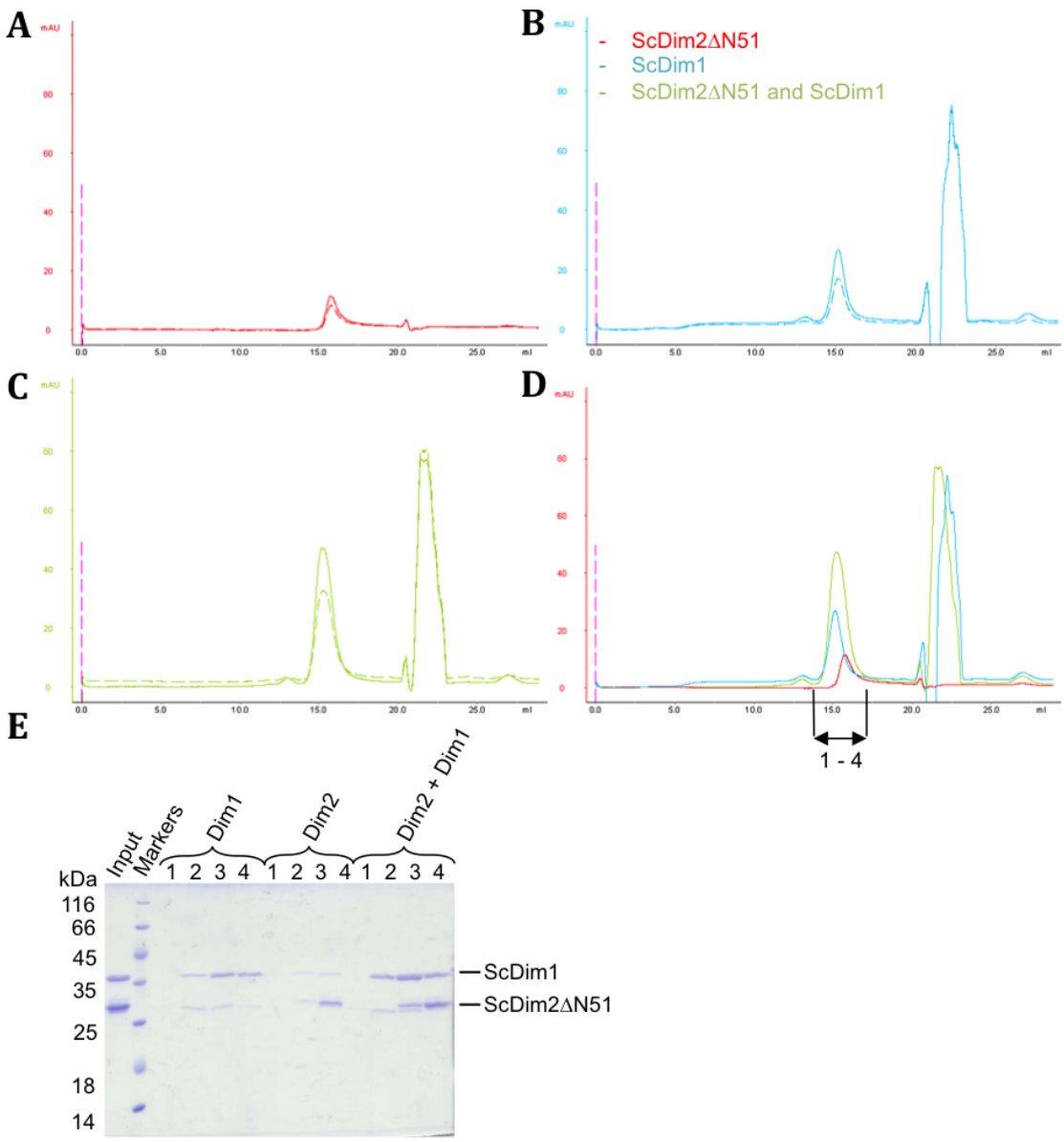


Figure 5.23: Semi-analytical SEC using ScDim2ΔN51 and ScDim1.
Absorbance at 280 nm is shown as a solid line whereas absorbance at 260 nm is shown as a dashed line. The key for the colours is shown on the right.

A) The SEC profile for ScDim2ΔN51 only.

B) The chromatogram for ScDim1 alone. Two peaks are seen. The more prominent one corresponds to the normal elution volume observed for ScDim1.

C) The SEC result when ScDim2ΔN51 is mixed with ScDim1 in a 1:1 ratio. It is impossible to say if the peaks remain separate as they overlap.

D) An overlay of the three previous results showing absorbance only at 280 nm. As the individual peaks overlap normally, it is difficult to say if a complex is formed. However, since the left edge remains in the same position, no increase in size is evident suggesting that there is no interaction.

E) Coomassie stained gel of the fractions indicated in panel **D** from each experiment. The distribution of the proteins does not change and they do not co-elute supporting that there is no complex formation.

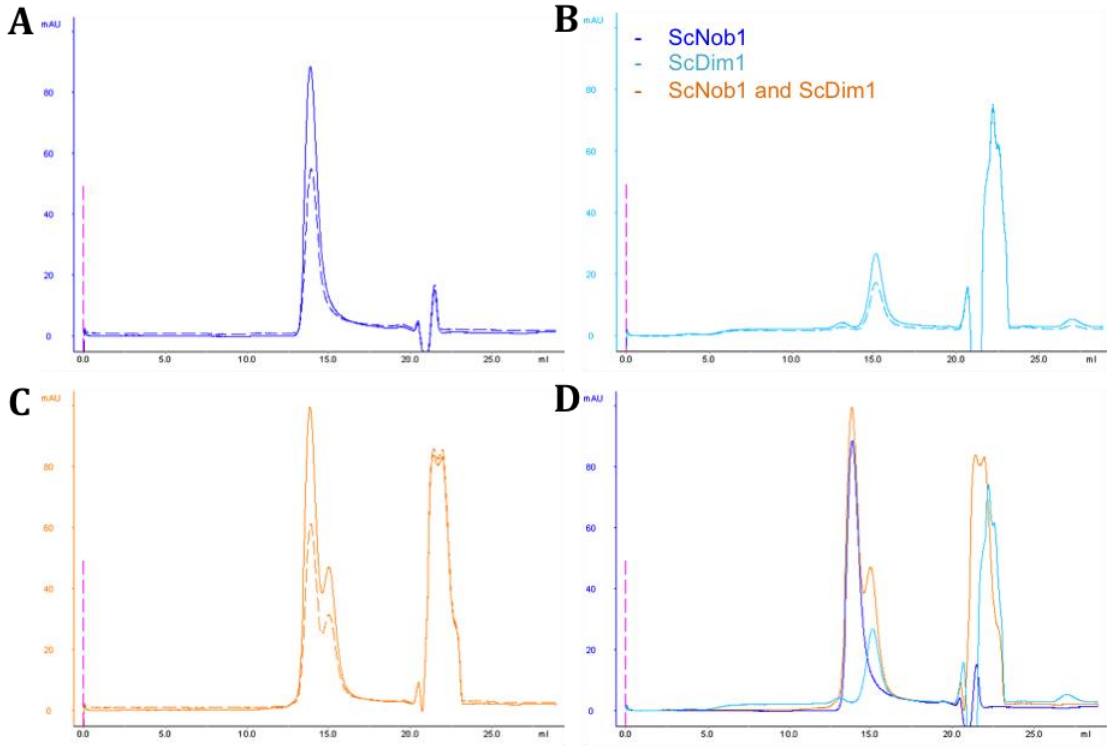


Figure 5.24: Semi-analytical SEC of ScNob1 and ScDim1.
Absorbance at 280 nm is shown as a solid line whereas absorbance at 260 nm is shown as a dashed line. The key for the colours is shown on the right.

A) The SEC profile for ScNob1 only. A single peak is seen.

B) The chromatogram for ScDim1 alone. Two peaks are seen but the one at the correct size is more prominent.

C) The SEC result when ScNob1 is mixed with ScDim1 in a 1:1 ratio. The two peaks remain separate.

D) An overlay of the three previous results showing absorbance only at 280 nm. The tops of the two peaks are in the same positions on the elution profile suggesting that there is no interaction. A slight increase in absorbance can be seen. This is likely due to pipetting or injection variation.

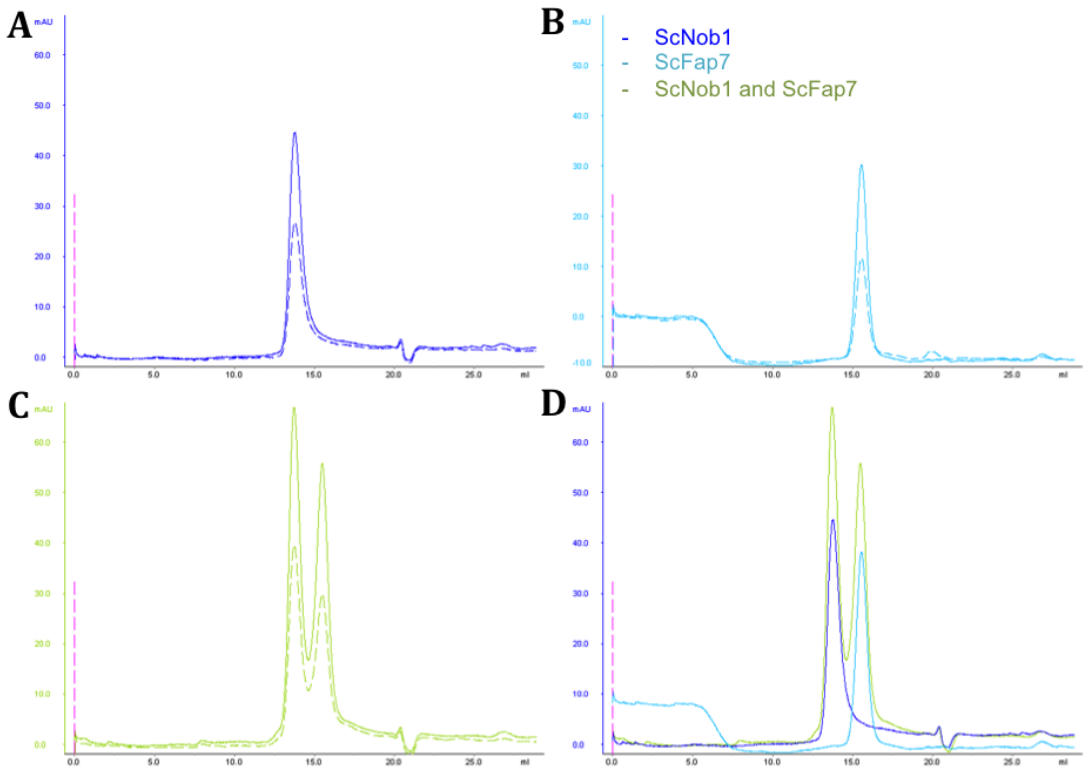


Figure 5.25: Semi-analytical SEC of ScNob1 and ScFap7.
Absorbance at 280 nm is shown as a solid line whereas absorbance at 260 nm is shown as a dashed line. The key for the colours is shown on the right.
A) The SEC profile for ScNob1 alone. Again, a single peak in the same position as before is seen.
B) ScFap7 elution profile. A single symmetric peak is seen suggesting the protein is clean.
C) SEC results for ScNob1 mixed with ScFap7 mixed in a 1:1 ratio. The two peaks remain separate and come out at the same elution volumes as they did separately. This suggests that there is no interaction between these two proteins in these conditions.
D) An overlay of the other three results showing absorbance only at 280 nm. Again, the peaks appear to stay in the same positions. The higher absorbance is likely due to variation in the pipetting or injection volume.

5.9 Dim2 Interacts with D-site Derived Sequences

5.9.1 Dim2 and the D-site tight hairpin do not interact

The Dim2 RNA binding site was predicted based on the archaeal structure in complex with bacterial RNA that co-precipitated during purification (GGAUCA sequence) (Jia et al, 2010). Different RNA sequences containing this site were transcribed as described for Nob1. They were tested for binding to the purified human and yeast Dim2 proteins. Initially, the RNA that forms the predicted hairpin

was used (DS0). When this sequence was mixed with ScDim2ΔN51, no interaction was seen in semi-analytical SEC (shown in **Figure 5.26**).

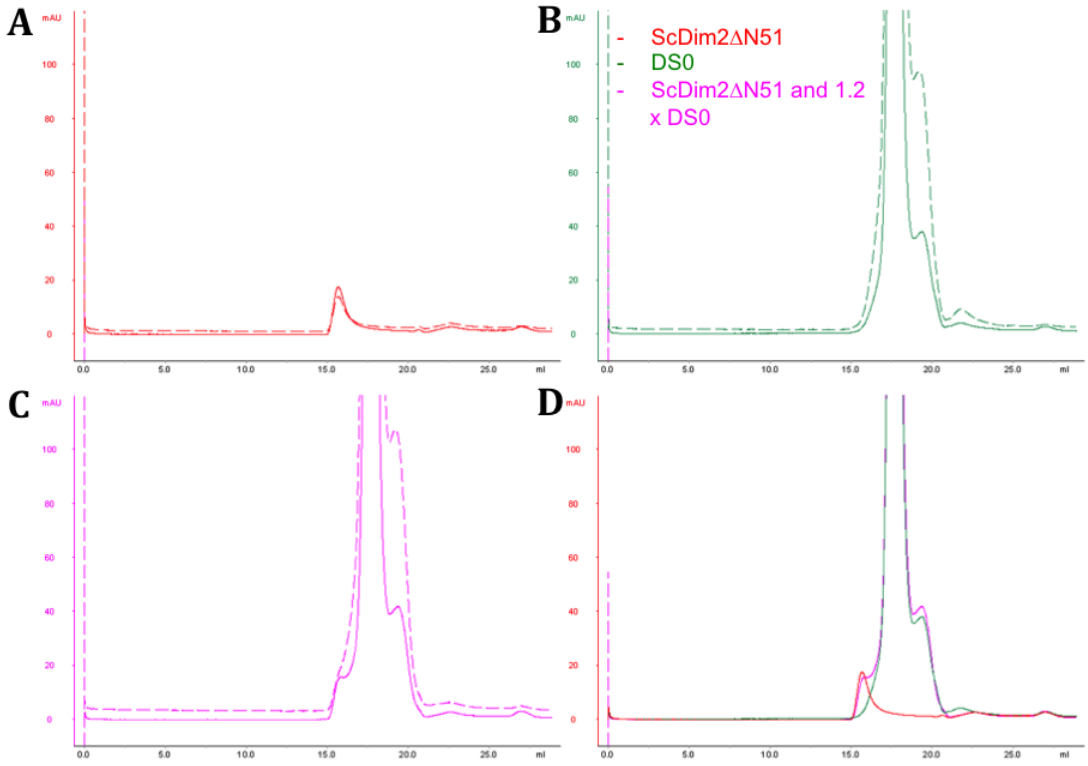


Figure 5.26: Semi-analytical SEC using ScDim2ΔN51 and yeast DS0 RNA.
Absorbance at 280 nm is shown as a solid line whereas absorbance at 260 nm is shown as a dashed line. The key for the colours is shown in the top right corner.

A) The SEC profile for ScDim2ΔN51 only.

B) The chromatogram for DS0 RNA alone. Two peaks are seen likely due to the formation of secondary structures at 4˚C.

C) The SEC result when ScDim2ΔN51 is mixed with DS0 in a 1:1.2 ratio. The peaks remain separate although an overlap is seen.

D) An overlay of the three previous results showing absorbance only at 280 nm. The peaks overlap to some extent. However, they are still visible as separate entities. Also, the absorbance remains the same. Thus, this protein and RNA do not make a complex in these conditions.

5.9.2 Dim2 and the D-site partial hairpin interact but they precipitate out of solution

Since the secondary structure of the D site remains in question, the semi-analytical SEC was repeated using the yeast DS1 RNA, which contains a shorter 3′ extension and does not form a stable hairpin. The GGAUCA sequence remains single-stranded in this RNA.

When this RNA was mixed with either, HsDim2 or ScDim2ΔN51, precipitation was observed in the tube. This was removed prior to injection onto the SEC column. **Figures 5.27** and **5.28** show the elution profiles for the two proteins, respectively. It appears that the precipitate contained each of the proteins and the RNA. This suggests that a complex is formed, but it is insoluble in these conditions.

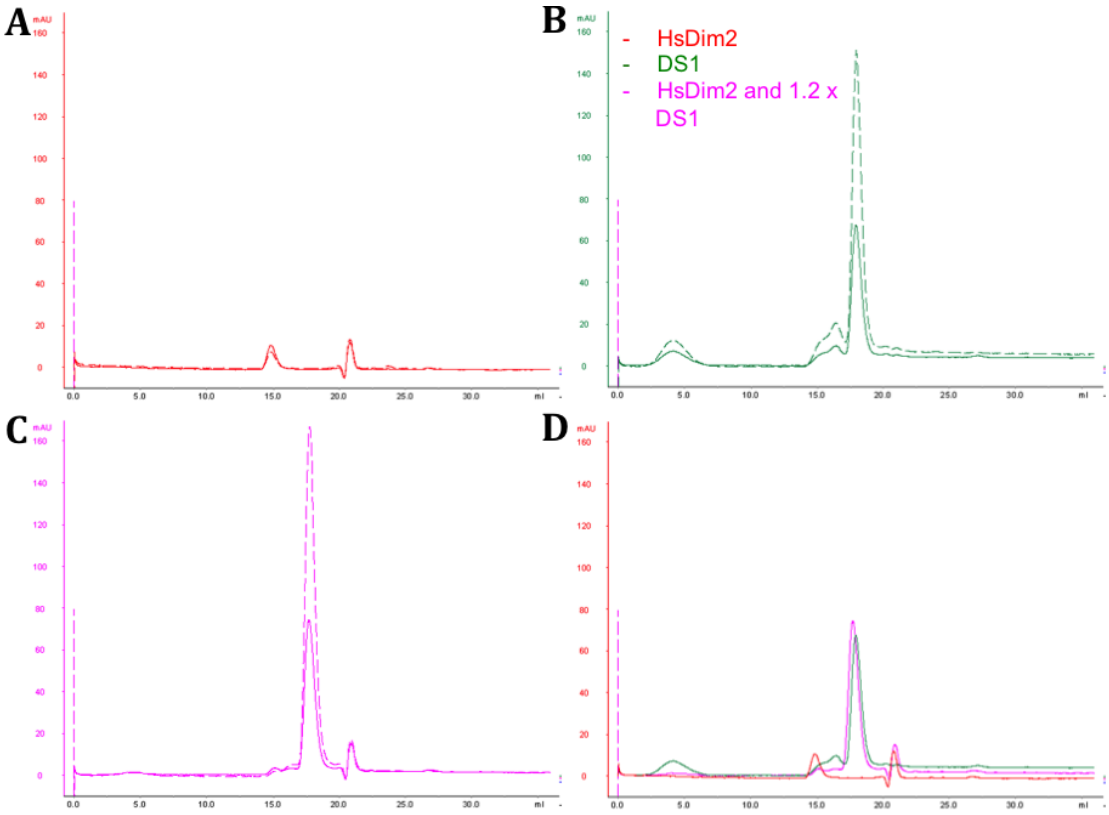


Figure 5.27: Semi-analytical SEC using HsDim2 and yeast DS1 RNA.
Absorbance at 280 nm is shown as a solid line whereas absorbance at 260 nm is shown as a dashed line. The key for the colours is shown on the right.

A) The SEC profile for HsDim2 only.

B) The chromatogram for DS1 alone. A couple of peaks are seen suggesting the formation of secondary structures. The samples were pre-incubated at room temperature to minimise these.

C) The SEC result when HsDim2 is mixed with DS1 RNA in a 1:1.2 ratio. Most of the protein peak disappeared when the precipitation was removed.

D) An overlay of the three previous results showing absorbance only at 280 nm. Almost all of the protein and some of the RNA were removed with the precipitation prior to the injection. This suggests that a complex was formed.

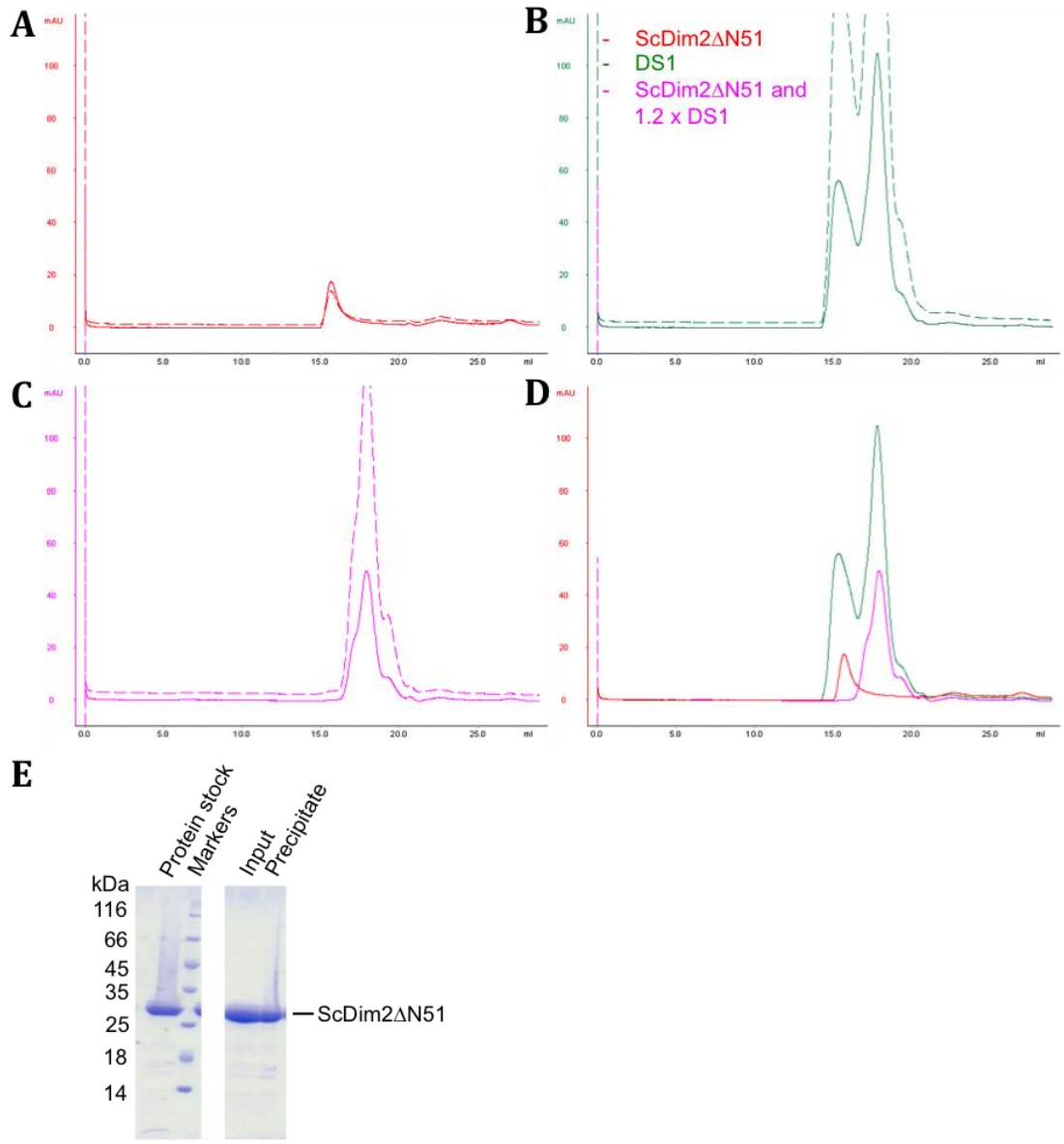


Figure 5.28: Semi-analytical SEC using ScDim2ΔN51 and yeast DS1 RNA.
Absorbance at 280 nm is shown as a solid line whereas absorbance at 260 nm is shown as a dashed line. The key for the colours is shown on the right.

A) The SEC profile for ScDim2ΔN51 only.

B) The chromatogram for DS1 alone. Two large peaks are seen suggesting the presence of several species of RNA. This is exacerbated by the pre-incubation being at 4˚C.

C) The SEC result when ScDim2ΔN51 is mixed with DS1 RNA in a 1:1.2 ratio. The precipitated particles were removed and the peaks that remain correspond to free RNA.

D) An overlay of the three previous results showing absorbance only at 280 nm. All of the protein appears to have been removed with the precipitation. A large amount of the RNA is also gone. This suggests complex formation.

E) Coomassie stained gel of the precipitated fraction. The precipitate definitely contained the protein. It is likely that it also contained RNA its level in the SEC elution profile was lower than expected. However, this was not confirmed.

An *in vitro* transcribed human DS1 RNA was also used with HsDim2. No interaction was observed (**Figure 5.29**); however, when the RNA was examined on a denaturing gel, multiple species of different lengths were seen (**Figure 5.29E**), suggesting that an incorrect fragment was synthesized.

In order to negate the precipitation issue, a solubility screen was performed. For this purpose, a drop of ScDim2ΔN51 was mixed with some DS1 RNA and different buffers. Precipitation was monitored under a microscope. It was found that the drops remained clear at higher pH (8.5 instead of 7.5) and higher salt (350 mM instead of 150 mM KCl). Thus, these two buffer conditions were used for further semi-analytical SEC. In both cases, the RNA was no longer able to interact with the protein (**Figures 5.30** and **5.31** respectively).

Finally, a Surface Plasmon Resonance (SPR) experiment was designed. In this assay, one of the binding partners is immobilised on a chip (ScDim2ΔN51) and the other one is run over the chip in solution (DS1 RNA). The chip changes mass when binding occurs. In turn, this causes a change in the angle of light refracted from the chip, allowing the kinetics of the interaction to be observed, including the association (when the RNA is added) and dissociation (when the RNA injection is stopped) curves. Also, the immobilisation of the protein could prevent precipitation when the interaction occurs.

The experiment involved five injections of increasing amounts of RNA and was performed at 4, 10, or 25˚C. Results at 4 and 10˚C were very similar but the higher temperature showed less noise. At 25˚C, very little response was seen (**Figure 5.32**). Two ways of analysing the results were tested. The kinetic analysis used the whole association and dissociation curves fitted to a one-to-one binding model. The other method (affinity analysis) averaged the last four seconds of each injection to obtain a curve of response vs. the concentration. However, this method requires the binding to be saturated.

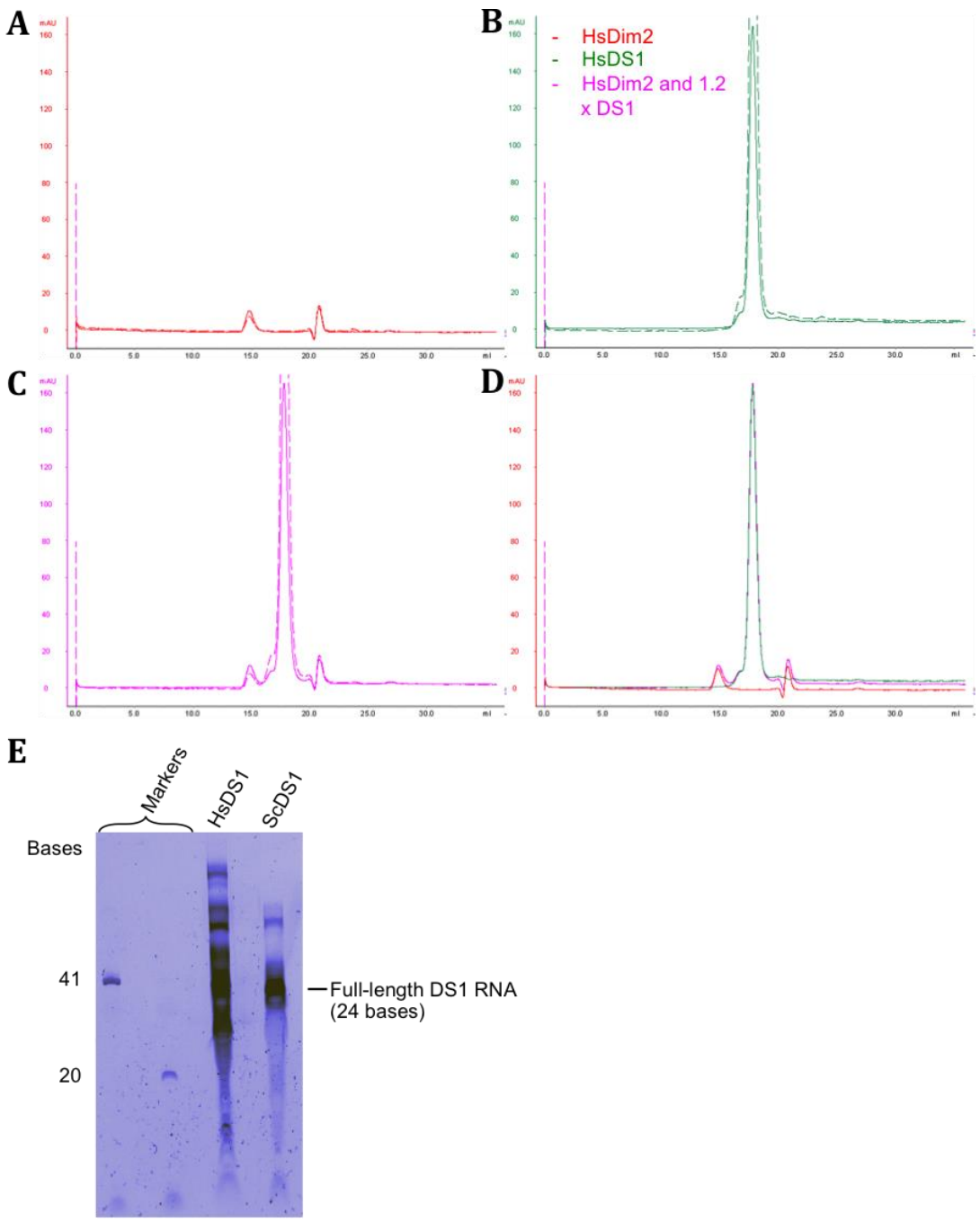


Figure 5.29: Semi-analytical SEC using HsDim2 and *in vitro* transcribed human DS1 RNA.
Absorbance at 280 nm is shown as a solid line whereas absorbance at 260 nm is shown as a dashed line. The key for the colours is shown on the right.

A) The SEC profile for ScDim2ΔN51 only.

B) The chromatogram for DS1 RNA alone. A single peak is seen.

C) The SEC result when HsDim2 is mixed with DS1 RNA in a 1:1.2 ratio. The two peaks remain separate.

D) An overlay of the three previous results showing absorbance only at 280 nm. Again, it is clear that the peaks remain separate suggesting that there is no interaction.

E) A denaturing gel of the *in vitro* transcribed yeast and human DS1 RNAs. The human sequence shows transcription of other fragments. Although the correctly sized fragment was gel purified, the sequence could still be incorrect.

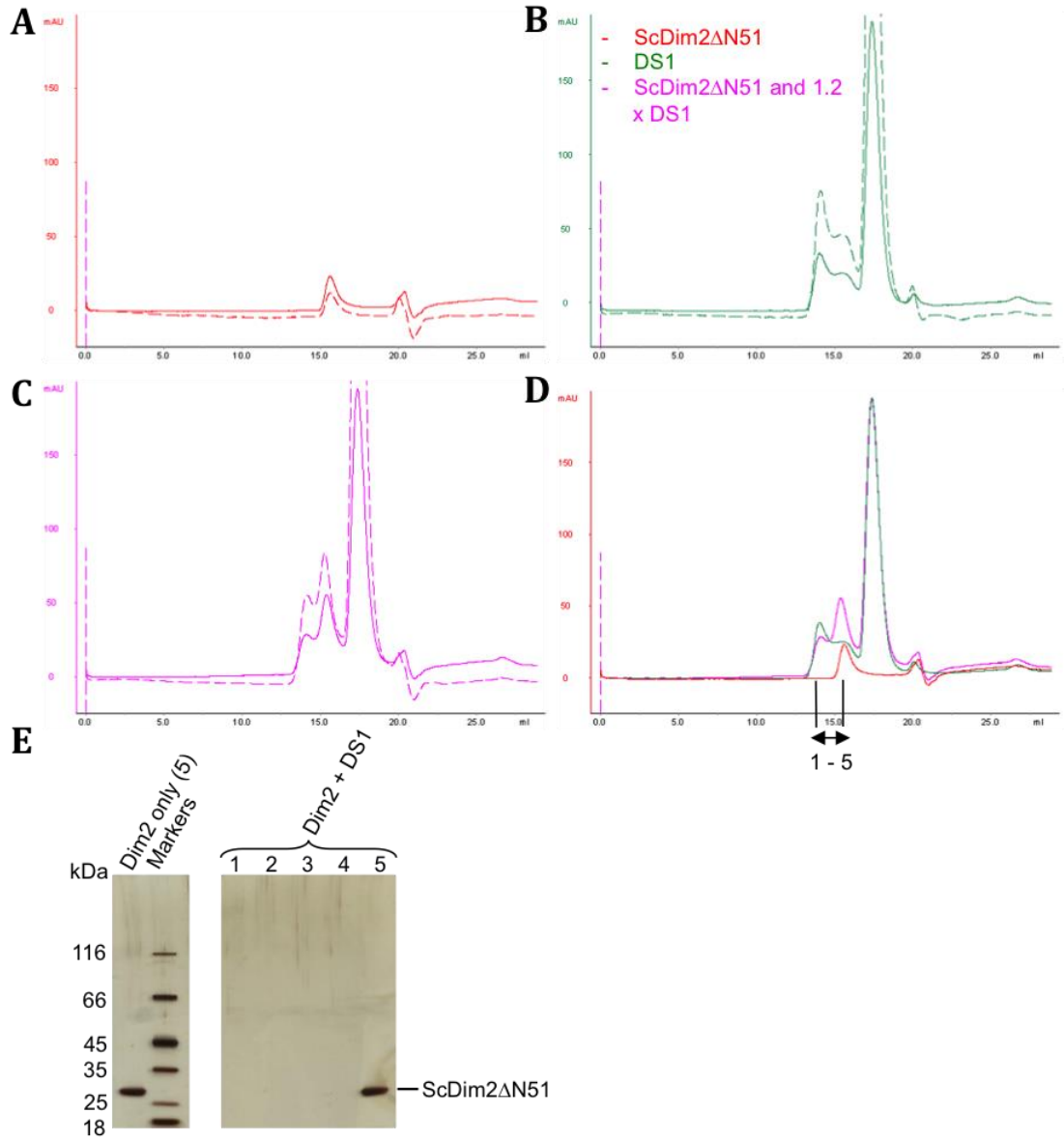


Figure 5.30: Semi-analytical SEC using ScDim2ΔN51 and yeast DS1 RNA at pH 8.5.
Absorbance at 280 nm is shown as a solid line whereas absorbance at 260 nm is shown as a dashed line. The key for the colours is shown on the right.

A) The SEC profile for ScDim2ΔN51 only.

B) The chromatogram for DS1 alone. Again, multiple peaks are seen, likely due to the incubation being at 4˚C.

C) The SEC result when ScDim2ΔN51 is mixed with DS1 RNA in a 1:1.2 ratio. Due to the complexity of the elution profile of the RNA alone, it is difficult to say if a complex is formed.

D) An overlay of the three previous results showing absorbance only at 280 nm. The peaks appear to be in the same positions; however, the absorbance does increase for some of them.

E) Silver stained gel of the fractions indicated in panel **D**. The distribution of the protein does not change (the last fraction for the Dim2-only experiment is shown in the left lane). Thus, there appears to be no complex formation in these conditions.

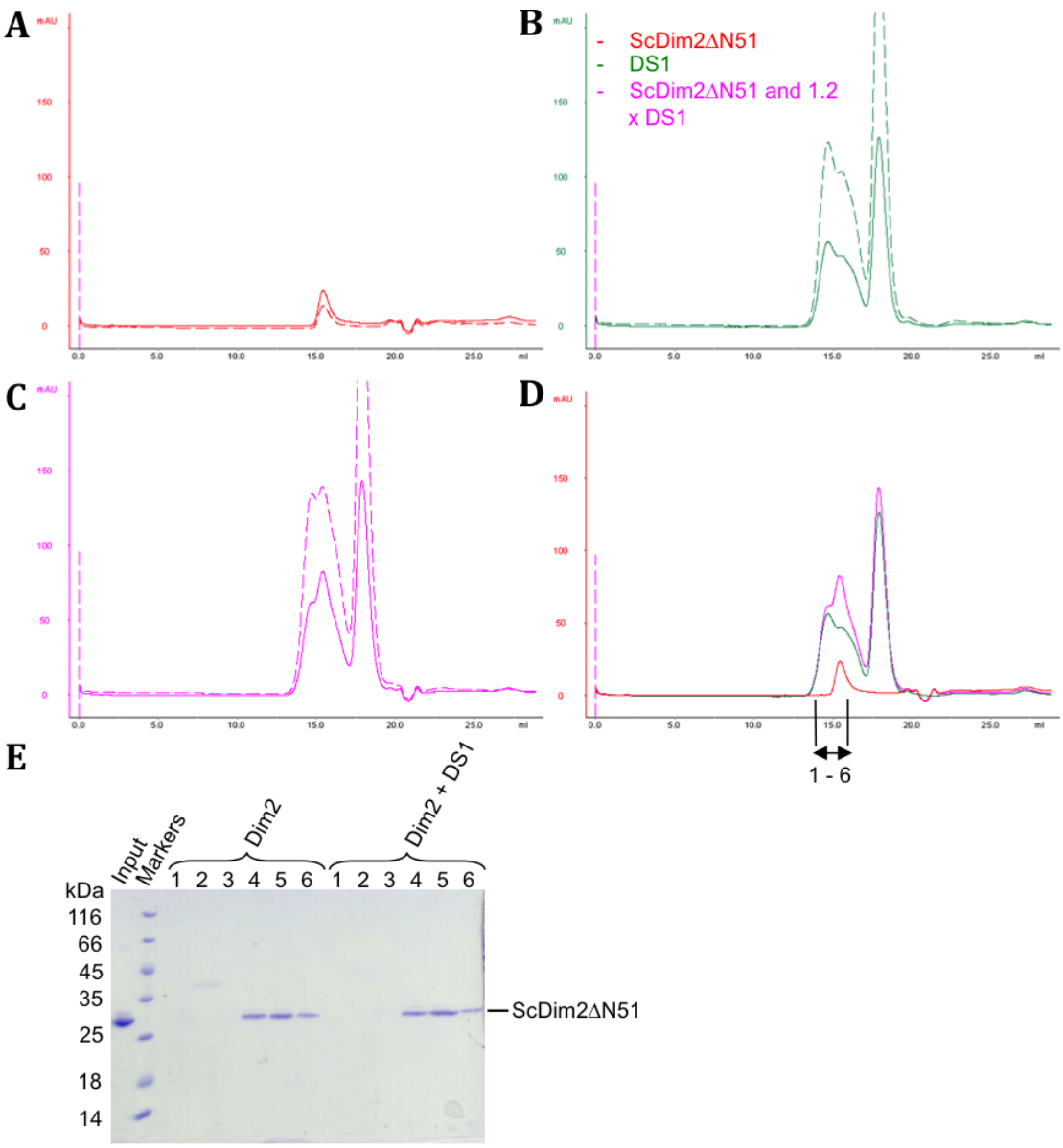


Figure 5.31: Semi-analytical SEC using ScDim2ΔN51 and yeast DS1 RNA in 350 mM KCl.
Absorbance at 280 nm is shown as a solid line whereas absorbance at 260 nm is shown as a dashed line. The key for the colours is shown on the right.

A) The SEC profile for ScDim2ΔN51 only.

B) The chromatogram for DS1 RNA alone. As before, multiple peaks are seen.

C) The SEC result when ScDim2ΔN51 is mixed with DS1 in a 1:1.2 ratio. It is impossible to say if the peaks remain separate as they overlap.

D) An overlay of the three previous results showing absorbance only at 280 nm. As the individual peaks overlap normally, it is impossible to definitely say if a complex is formed. However, since the left edge remains in the same position, no increase in size is evident, suggesting that there is no interaction.

E) Coomassie stained gel of the fractions indicated in panel **D** from each experiment. The distribution of the protein does not change supporting that there is no complex formation.

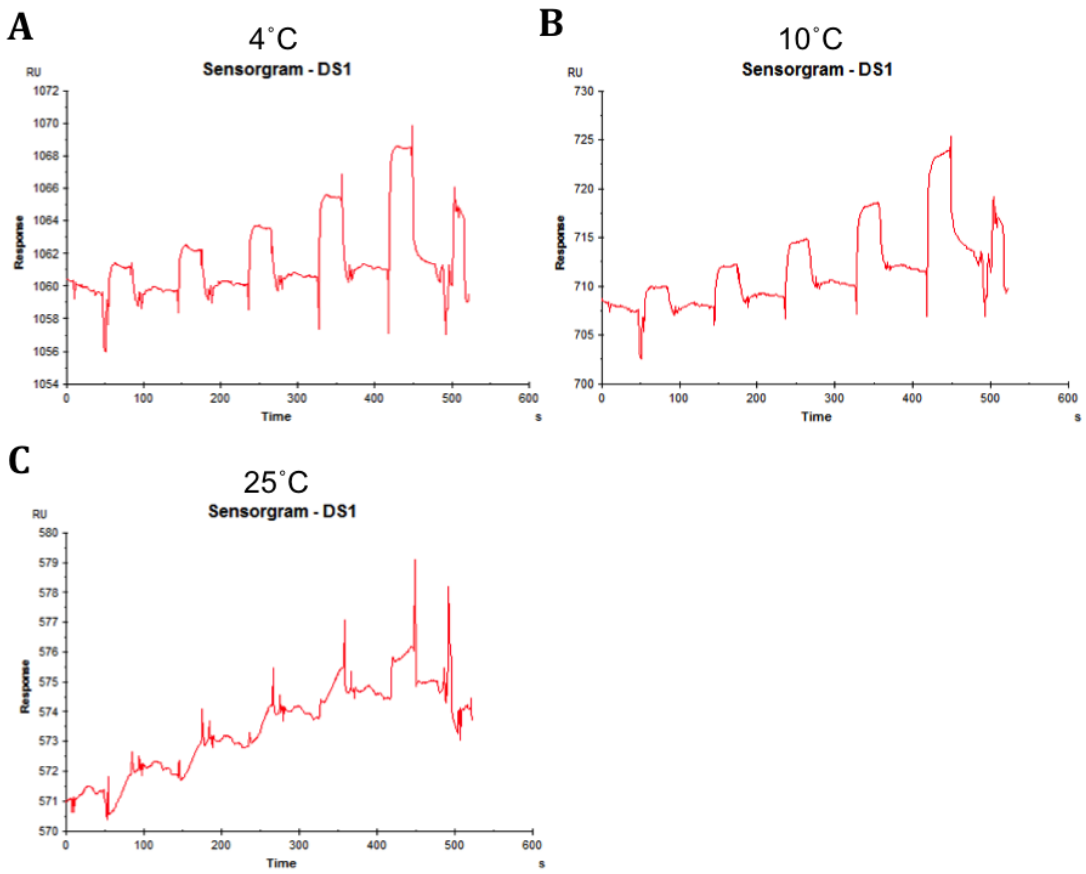


Figure 5.32: The raw sensogram results obtained from the SPR experiments at different temperatures.

The starting response corresponds to the nickel and protein loaded onto the chip. Each experiment involved five injections of increasing concentrations of DS1 RNA. 15 µl of each concentration were injected followed by a 30 second break. The overall response (R_{max}) can be seen in **Table 5.3**.

A) The sensogram obtained at 4˚C. The RNA was seen to bind (the rise in response). After the injection, there was a fast dissociation (the sudden drop in response) followed by a slower one (gradual drop). The binding is seen to be concentration dependent.

B) The results at 10˚C. The sensogram looks very similar to the one obtained at 4˚C. However, fewer buffer mismatches are seen (the vertical lines at the start and end of each injection).

C) The sensogram obtained with the chip at 25˚C. Some binding is still present and the baseline is raised with each injection. However, the association and dissociation curves are difficult to see due to noise. Thus, it appears that the most clearly interpretable data was obtained at 10˚C.

Kinetic analysis results for the 10˚C experiment, shown in **Figure 5.33**, suggest that an interaction is present. However, the scale of the response seen is much lower than predicted. **Table 5.3** shows the predicted maximum response and the values obtained using the assay at different temperatures. Large bulk contributions were observed in the kinetic analysis, which suggests that precipitation might still be happening on the surface of the chip. This is supported by the nature of

the dissociation curve, which shows two different 'off' rates. The faster rate would correspond to larger molecules (i.e. precipitate) coming off, while the non-precipitated individual RNAs dissociating would cause the slower rate.

The affinity analysis was also performed (**Figure 3.33C**). However, the association curves are not flat at the top showing that the chip was not saturated. Thus, the values obtained from this analysis are underestimated and the dissociation constants could not be trusted. Overall, a very small portion of the surface appears to be active and the RNA precipitates onto it.

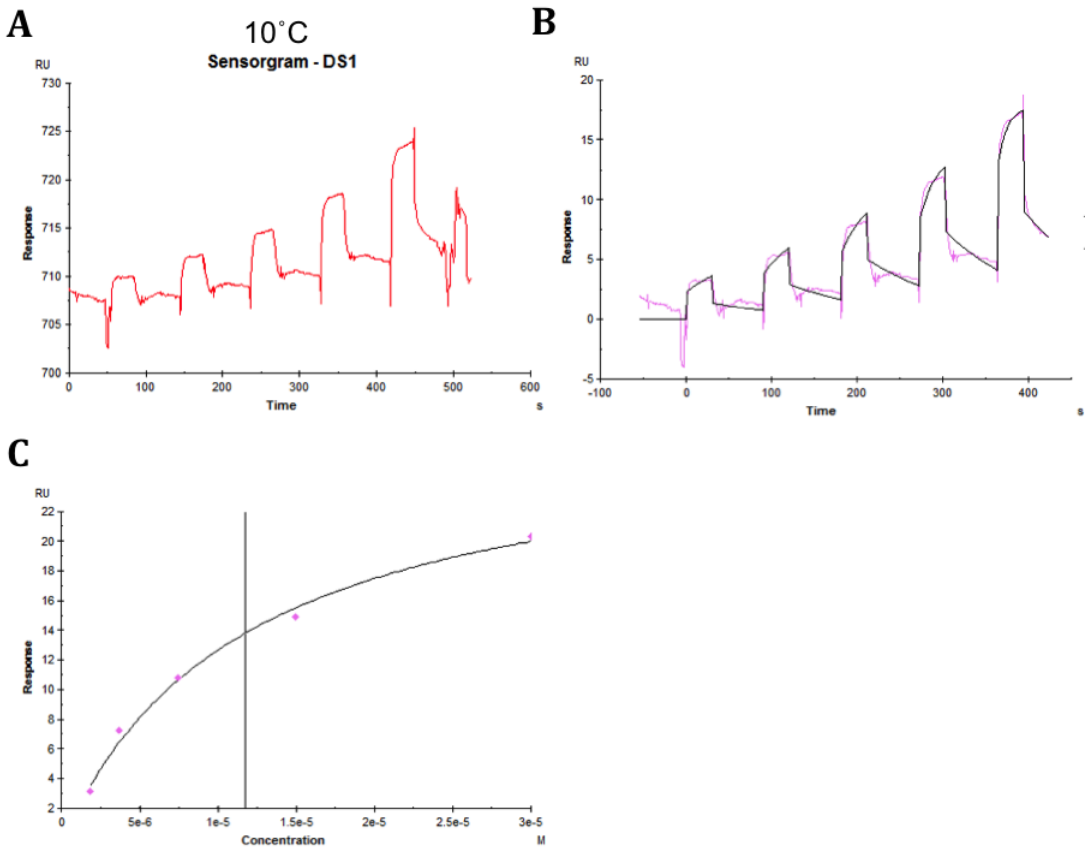


Figure 5.33: The analysis results for the SPR experiment performed at 10˚C.

A) The sensogram is shown again for comparison.

B) Kinetic analysis of the sensogram data. In pink, the raw data is shown, while the model fitted is shown in black. This corresponds to a one-to-one binding model. It does not fit the data as there are two on and off rates observed (fast and slow).

C) Steady state affinity analysis of the sensogram. The last four seconds of each injection were averaged for this purpose. Thus, the response could be plotted relative to the DS1 concentration. However, as the binding was not saturated (i.e. the top of each peak is not flat) the results are inaccurate.

Table 5.3: Table of the predicted response for DS1 binding to a chip loaded with ScDim2ΔN51. The results for each of the three temperatures are shown. The second column lists the response units (RU) obtained just from loading the protein onto the chip (it corresponds to the initial RU seen minus the RU obtained from loading the chip with nickel). Then, the predicted R_{max} is shown. This is the maximum response obtained if all of the surface were to bind the RNA. It was calculated using the formula: R_{max} = (MW of RNA)/(MW of protein) x $RU_{protein}$. Finally, the R_{max} obtained using the two different analysis methods is shown.

Temperature (˚C)	Protein loaded (RU)	Predicted R_{max} (RU)	Kinetic Analysis R_{max} (RU)	Affinity Analysis R_{max} (RU)
4	597	170	5	15
10	587	167	10	28
25	463	132	3	3

5.9.3 ScNob1 and HsDim2 compete for the D site RNA

Since it appears that Dim2 and Nob1 interact with overlapping RNA sequences, a semi-analytical SEC experiment was performed using both proteins and DS1 RNA. It was found that the two proteins (ScNob1 and HsDim2) appear to be competing for the nucleic acid (**Figure 5.34**).

After all the reactants were mixed, precipitation was observed (as for HsDim2 and RNA only). The tube was spun down and the precipitation removed prior to injection onto the SEC column. The elution profile showed a peak at the same volume as for ScNob1 and RNA alone. This suggests that both complexes were present in the tube, independently. There was no evidence that a larger complex containing all three components was formed. In the future, a silver stained gel of the fractions should be included in the analysis to confirm this.

The lack of an interaction between these two proteins was previously observed. However, if they both could bind to the RNA as the same time, a larger complex would still be formed. Due to the precipitation issue and because the yeast Dim2 protein was obtained in the late stages of the project, the experiment was not repeated with ScDim2ΔN51. This could be done in the future but, due to the precipitation, the results are likely to be the same as with HsDim2. Also, the buffer conditions need to keep the pH at 7.5 and salt at 150 mM to preserve the RNA-Dim2 interaction. The protein-protein interactions would be too weak in these conditions to be observed (as seen previously). Thus, it appears that the two proteins cannot interact with the RNA at the same time.

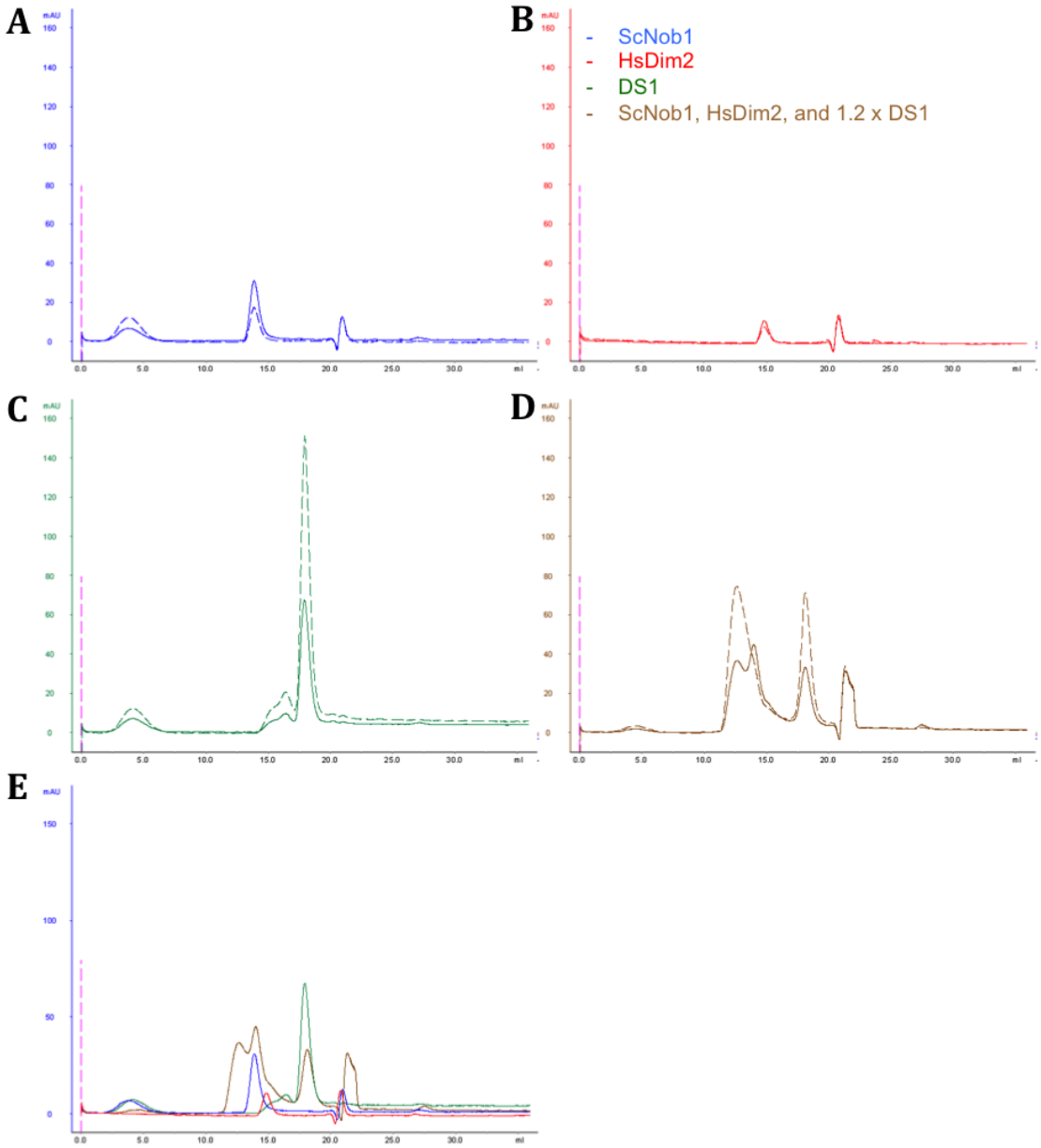


Figure 5.34: Semi-analytical SEC using HsDim2, ScNob1, and the yeast DS1 RNA.
Absorbance at 280 nm is shown as a solid line whereas absorbance at 260 nm is shown as a dashed line. The key for the colours is shown on the right.

A) The elution profile for ScNob1 alone.

B) The SEC profile for HsDim2 only.

C) The chromatogram for DS1 alone. As before, multiple peaks are seen. This is minimised at room temperature.

D) The SEC result when HsDim2 is mixed with ScNob1 and DS1 RNA in a 1:1:1.2 ratio. Precipitation was present and removed prior to the injection. The peaks correspond to ScNob1 alone and the ScNob1-DS1 complex.

E) An overlay of the three previous results showing absorbance only at 280 nm. Since ScNob1 remained in solution and HsDim2 precipitated, it appears that both proteins bind to the RNA independently and compete for the site. A silver stained gel of the fractions was not run but should be included in the future to confirm the results.

5.10 Crystallisation Trials

Human and truncated yeast Dim2 (ΔN51) without RNA were tested in crystallisation screens. However, no hits were obtained. The majority of drops showed some granular precipitation. Screening with RNA-protein complexes were not performed due to the precipitation issue. In the future, a shorter RNA fragment should be tested (e.g. just the GGAUCA sequence) to see if the precipitation can be controlled.

The proteolysed fragments were also tested for crystallisation. However, these fragments also produced no hits. The predicted N-terminal truncation could be used to re-clone the HsDim2 without these sections. The protein could then be purified with fewer contaminants for crystallisation.

5.11 Discussion

5.11.1 ScDim2 modelling based on the archaeal structure

The modelling performed using the archaeal structure showed the predicted fold for the KH and KH-like domains. However, the validation of the models showed a significant number of outliers. More outliers were present in the degenerated domain supporting the idea that it is unable to bind to RNA in the same manner. This is further reinforced, as the RNA binding site for this domain is not conserved between archaea and eukaryotes (unlike the site for KH2). The GxxG motif of the functional KH domain could form the same interactions with the RNA as in the archaeal protein as no geometry prohibitions were found. The model provides a good starting point but a structure of a eukaryotic protein is still highly desirable, especially to understand the structure and function of the first, non-canonical KH domain.

5.11.2 Soluble Dim2 was purified

Many different constructs of Dim2 were tested for expression and purification. It was found by Iva Tchasovnikarova that the full-length yeast protein

eluted with the bacterial ribosome and the purification of this construct was not successful during the course of this project. However, directions for future attempts have been uncovered. It was found that truncated yeast Dim2 cannot interact with RNA at high pH or salt concentrations. Thus, future purification methods could incorporate these conditions to try to separate the full-length protein from the ribosome.

N-terminally truncated versions of yeast and human Dim2 were cloned, as this portion of the protein was the least conserved (see **Figure 1.7**). When ScDim2 was truncated at residue 91 or 100, no soluble protein was obtained. However, truncations at residues 51 and 61 showed good overexpression in the small-scale trials. The equivalent deletions in the human protein (ΔN31 and 42) also showed soluble expression.

ScDim2ΔN51 was then expressed in large scale and successfully purified. This construct was His-tagged and, thus, purified at pH 8.0 unlike the GST-tagged full-length protein (purification attempts at pH 7.5). The higher pH could have prevented the interactions with the bacterial ribosome. The full-length GST-tagged human protein, which has a shorter N-terminal extension, was also successfully purified. Unlike the full-length yeast protein, HsDim2 did not co-purify with the bacterial ribosome. It is possible that it forms fewer interactions with the bacterial rRNA due to sequence divergence between the two organisms (the two proteins show 48.2% sequence identity).

Finally, the protein from *C. thermophilum* was tested in expression trials and found to be insoluble. Again, N-terminal truncations could have potentially helped. However, no further tests with this protein were performed.

5.11.3 Recombinant Dim2 precipitates upon interaction with RNA

Both, the full-length human and the truncated yeast proteins, showed precipitation upon interaction with the D site RNA. Thus, co-crystallisation trials could not be performed. A solubility screen revealed the lack of precipitation at higher pH and higher salt concentrations. However, this was caused by the lack of complex formation and not solubilisation of the complex. Nevertheless, this result does support that a complex is formed at physiological conditions.

The full-length yeast Dim2 could not be easily purified since it co-purified with bacterial RNA and ribosomal proteins. However, when it was incubated with an RNase, it precipitated. It is possible that the protein bound to RNA equivalent to the eukaryotic D site sequence and protected it from digestion. Then, the protein precipitated in the same manner as the N-terminal truncation mixed with pre-made RNA.

In order to overcome the solubility issue, ScDim2ΔN51 was immobilised prior to addition of RNA in an SPR experiment. However, there likely still was precipitation on the chip surface as two on and off rates were observed. Also, the surface showed a low percentage of activity. This could have been caused by precipitation preventing parts of the chip interacting with the RNA or bacterial RNA contamination co-purified with the protein. The latter explanation is less likely as the absorbance at 260 nm did not show significant RNA contamination during purification.

In the future, shorter RNA sequences (e.g. the GGAUCA sequence only) could be used to test for the formation of a soluble complex. Also, if the RNA was biotinylated, it could be coupled to the SPR chip (via streptavidin) and the protein could be run over in the solvent channels. If the RNA was relatively sparse on the chip, this could prevent precipitation onto the surface.

Interestingly, ScDim2ΔN51 interacted with the partially folded D site (DS1) but not the tight hairpin (DS0), similar to Nob1. This, again, supports the idea that the D site is normally single-stranded or that another factor is involved in unwinding it to allow binding.

The human Dim2 protein was able to interact with the yeast DS1 but not the *in vitro* transcribed human DS1 sequence. However, this is likely due to errors in transcription as the gel showed the presence of many different RNA species. Thus, for other experiments using this sequence (including the binding assay with ScNob1 described earlier), the RNA was ordered from IDT as an HPLC purified fragment.

5.11.4 The oligomeric states of Dim2 by itself and in complex remain unknown

SEC data did not provide an accurate determination of the oligomeric state of human or yeast Dim2. The proteins appear to be larger than a monomer in SEC.

However, as mentioned earlier, the elution volume depends on shape and interaction with the matrix as well as the size. HsDim2 should be about 3 kDa larger than ScDim2ΔN51 and eluted about 2.5 ml earlier in SEC. This suggests that the two constructs are a similar shape/size and the N-terminal truncation did not have a significant effect on conformation.

Shape prediction using the archaeal structure and HYDROPRO software shows that Dim2 is marginally elongated. However, the elongation would not be enough to explain the early elution volumes. It is likely that a more globular dimer is formed. Otherwise, the early elution could be caused by proteins interacting differently with the SEC matrix.

SEC-MALS followed by DLS should be performed on the different proteins in order to obtain accurate molecular weights and information about the shapes of the different species. If a soluble complex between Dim2 and RNA had been obtained, it could also have been tested in this manner. Semi-analytical SEC would have been performed with differing ratios of protein and RNA (not just 1:1.2 or 1:4). An excess of protein would also have been tested in case more than one Dim2 molecule is able to bind to each RNA particle. However, due to the solubility problems none of these experiments were performed.

5.11.5 Nob1 and Dim2 compete for binding to the D site

The experimentally determined D site RNA binding sequence for Nob1 (AGGA) appears to overlap with the predicted Dim2 binding site (GGAUCA). Although the Dim2 site was predicted from an archaeal structure and not experimentally determined, it is conserved from archaea to mammals. Also, the Dim2 domain predicted to interact with this site (KH2) is more highly conserved than other parts of the protein. Thus, it is likely that eukaryotic Dim2 would interact with this site in a similar manner.

Since the RNA sites for Nob1 and Dim2 overlap, it is likely that the two proteins cannot interact with the D site at the same time. This would explain why Nob1 is unable to cleave the D site until late stages of 20S pre-rRNA maturation. Dim2 might have to be released before Nob1 can perform its function making sure it occurs at the right time in the pathway. Semi-analytical SEC results using all three

components showed no appearance of a larger complex suggesting that it is not formed. In this experiment, the human Dim2 protein was used with yeast Nob1. These proteins showed no interactions with each other in the conditions used. It is possible that protein-protein interactions are needed to create a stable complex. However, they would also mask whether competition for the D site RNA exists (i.e. a larger complex could be mediated by protein-protein interactions only and not by both proteins binding to the same RNA fragment). Using the two proteins from different organisms supports that the competition does exist. Nevertheless, these results are inconclusive due to the precipitation of HsDim2 with RNA and further tests are needed.

Previous studies and new cross-linking data have shown that Nob1 and Dim2 interact with each other as well as the D site RNA. As mentioned above, when human Dim2 and yeast Nob1 were mixed and analysed using semi-analytical SEC, no change in absorbance or size was seen. This could be due to the small differences in sequence between the yeast and human versions of the proteins. Also, the interaction could be weak without a long RNA scaffold, incorporating the ITS1 sequence for Nob1 binding and the D site for Dim2 interactions. Finally, the buffer conditions also likely played a part and higher salt and pH should be tested in the future to confirm this result.

When the two yeast proteins are mixed, a shift in elution volume is seen and SDS-PAGE analysis of the fractions confirms that Nob1 and Dim2 co-elute. However, this only occurred at higher pH and salt concentrations. The profile for the Dim2-only peak is still seen suggesting that the free protein is still available. This could be caused by the oligomerisation states of the two components and their relative ratios. For example, a Nob1 dimer could be binding to a Dim2 monomer. SEC-MALS on the complexes would have to be performed in order to obtain accurate molecular weights within each peak. Then, the composition would be calculated based on the theoretical sizes of the components. The extra free Dim2 protein would also be present if the interaction is weak without an RNA mediator.

Alternatively, if the N-terminus of the Dim2 protein is involved in the protein-protein interaction, it could be weakened when ScDim2ΔN51 is used instead of the full-length protein. This is unlikely, as this portion of the protein is not

conserved. Finally, if any post-transcriptional modifications were involved, these would be missing in the recombinant proteins. For example, phosphorylation might be required to separate dimers into monomers, which can then form heterodimers.

In the future, an SPR experiment containing ScNob1, ScDim2ΔN51, and D site RNA could be designed. As mentioned earlier, the RNA could be coupled to the surface. Then, each of the proteins could be washed over the surface. Competition for binding could be observed. If both proteins were able to bind concomitantly, this would also be seen.

5.11.6 Previously observed interactions could not be replicated

A recent NMR study has shown that archaeal Nob1 and Fap7 interact with each other (Hellmich et al, 2013). An attempt was made to replicate this result with the yeast recombinant proteins in a semi-analytical SEC experiment. The two proteins did not interact in these conditions. This discrepancy could be explained by the increased complexity of the ribosome biogenesis in eukaryotes and the eukaryotic protein sequences. It is possible that, in eukaryotes, the two proteins are in close proximity and that the interaction is mediated by other proteins or RNA and is very weak without the extra components.

The same result was obtained when the ScDim2ΔN51 was tested for an interaction with ScDim1. This has also been observed in the past (Vanrobays et al, 2004). However, no binding was observed with the recombinant proteins. Again, it is likely that other components of the ribosome biogenesis pathway are needed to mediate the interaction. Otherwise, the N-terminus of Dim2 might be involved in complex formation. This is less likely as it is not conserved in sequence between organisms but cannot be ruled out at this stage.

An experiment testing whether Nob1 interacts with Dim1 was also performed. There is no previous evidence for a direct interaction between these two proteins. However, the negative result acts as a good control for other semi-analytical SEC experiments. Comparison between Nob1-Fap7, Dim2-Dim1 and Nob1-Dim1 results clearly shows that none of these interactions are observed in the conditions tested.

5.11.7 Crystallisation trials of Dim2 have been unsuccessful to date

Crystallisation screens using either ScDim2ΔN51 or HsDim2 were unsuccessful. In the future, further screens could be performed. The proteolysed HsDim2 fragment appeared to be an N-terminal truncation. The corresponding truncation could be cloned, expressed, and purified to remove the variability when using proteases. Then, it could be re-screened for crystallisation. Also, multiple constructs cloned during the course of the project have not yet been purified. These proteins could be tested in crystallisation screens in the future.

Co-crystallisation of Dim2 and Nob1 has not yet been performed. However, the two proteins could be co-purified in order to obtain a stable complex that could then be used for crystallisation. Buffer conditions that increase the amount of complex formed have been already identified (higher pH and salt) and could be further optimised.

5.12 Concluding remarks

Dim2 constructs were successfully purified and were shown to interact with single-stranded RNA. However, the proteins precipitated upon interaction with RNA and this could not be prevented during the project. It appears that Dim2 and Nob1 compete for the D site RNA during ribosome biogenesis. At the same time, Nob1 and Dim2 form protein-protein interactions with each other. Thus, this could form a regulatory scaffold that ensures that the D site cleavage occurs in the right place at the right time. In this case, Dim2 could recruit Nob1 to the ribosome. Then, Dim2 would have to leave for Nob1 to bind and cleave the D site.

6 OPTIMISATION OF TSR1 CRYSTALS

6.1 Introduction

Tsr1 is another ribosome biogenesis factor, which is needed for the late stages of small subunit maturation. It shows sequence similarity to several GTP-binding proteins (e.g. Bms1) but lacks the regions identified to be involved in the actual binding of GTP (Gelperin et al, 2001). Its interactions with rRNA have been mapped to a site similar to that bound by a bacterial GTPase, Era (Granneman et al, 2010).

Although the exact function of Tsr1 remains unknown to date, it is predicted to prevent premature assembly of the ribosomal subunits and translation initiation. It has been suggested to be involved in the nuclear export of the pre-40S particle as well (Campbell & Karbstein, 2011; Gelperin et al, 2001; Schafer et al, 2003; Strunk et al, 2011). Its positioning near the D site could also prevent premature cleavage by Nob1 (Granneman et al, 2010) while making sure that the 20S pre-rRNA does not get degraded before this cleavage occurs (Strunk et al, 2012).

In order to better understand Tsr1 function, a structure of the eukaryotic protein is needed. This chapter reports the initial crystallisation trials for the yeast protein. They were successful although, due to time constraints, the full structure has not yet been solved.

6.2 Native Crystal Optimisation

Dr. Uma Jayachandran designed the constructs, purified and characterised them, and set up the initial commercial crystallisation screens. Initially, the full-length protein was purified. However, it did not yield any hits in crystallisation screens. Subsequently, it was subjected to limited proteolysis. The assay showed the presence of two stable bands. When the assay was scaled up and the products were analysed by SEC, it was found that both fragments co-purified together. MALDI-ToF analysis identified the two fragments as parts from the N- and C-termini.

However, the middle portion of the protein was missing. Thus, different constructs missing the middle loop were created and characterised (like the loop-out mutants for Nob1). The two domains remained connected but by a shorter linker in these constructs.

These new constructs were also cloned, purified, characterised, and screened for crystallisation by Dr. Uma Jayachandran. Multiple hits were obtained with ScTsr1Δ410-476. The three most promising conditions observed are shown in **Table 6.1**.

Table 6.1: The three most promising hits obtained using ScTsr1Δ410-476

Condition	Buffer	Salt	Precipitant
1	50 mM MES pH 6.0	0.2 M Sodium malonate	10% PEG 8000
2	50 mM MES pH 6.0	0.2 M Magnesium chloride	10% PEG 3350
3	50 mM MES pH 6.0	0.2 M Calcium acetate	18% PEG 8000

I further optimised all three conditions in larger 24-well sitting drop plates. Only condition 1 showed large crystals. Thus, this condition was chosen for further optimisation. Buffers at various pH values were tested and it was found that Tris-HCl at pH 8.0 gave more evenly shaped crystals (**Figure 6.1**). However, these needles stuck to the bottom of the plate and dissolved over time.

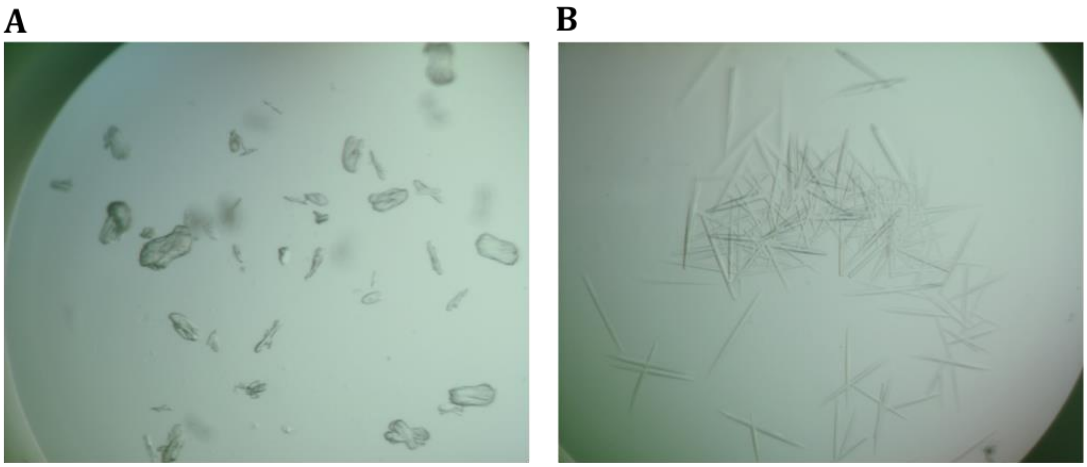


Figure 6.1: Sample crystals obtained using ScTsr1Δ410-476.

A) Crystals obtained using the original condition 1 at pH 6.0. Large barrel shaped crystals (about 210 µm in the longest direction) were observed. However, multiple crystals were stuck together and it was impossible to separate them.

B) Crystals obtained when the pH was raised to 8.0. Single needles, about 330 µm in length, were observed. However, they stuck to the bottom of the plate and dissolved over time.

Nevertheless, the needles were harvested and sent to the Diamond Light Source synchrotron in Oxford and tested for diffraction. For this purpose, 4 images separated by a 45˚ rotation were collected for each crystal. The best diffraction obtained was very weak and extended to about 20 Å in resolution. This was not improved when using hanging drops instead of sitting drops (to prevent the crystals sticking to the wells). Streak seeding was also performed. However, the diffraction did not improve. Finally, additive screens were also undertaken but these showed no improvement in crystal appearance or diffraction.

Since the N-terminus of Tsr1 is also predicted to be unstructured, further constructs without this sequence were designed and tested for crystallisation in commercial screens by Dr. Uma Jayachandran. It was found that ScTsr1ΔN46Δ410-476 (**Figure 6.2**) crystallised in these screens. Conditions giving crystals with this construct are listed in **Table 6.2**.

Table 6.2: The hits obtained using ScTsr1ΔN46Δ410-476.

Condition	Buffer	Salt	Precipitant
1	50 mM MES pH 6.0	0.2 M Sodium malonate	10% PEG 8000
2	50 mM MES pH 6.0	0.08 M Ammonium sulfate	4% MPD
3	50 mM MES pH 6.0	0.2 M Ammonium acetate	4% MPD
4	50 mM MES pH 6.0	0.2 M Ammonium sulfate	10% PEG 8000
5	50 mM MES pH 6.0	0.2 M Ammonium sulfate	18% PEG 8000
6	50 mM MES pH 6.0	0.2 M Sodium chloride	18% PEG 8000
7	100 mM Bis-Tris pH 6.5	0.2 M Ammonium acetate	25% PEG 3350
8	-	0.2 M Ammonium tartrate dibasic pH 6.6	20% PEG 3350
10	-	0.1 M Sodium malonate pH 6.0	12% PEG 3350
11	-	0.2 M Ammonium tartrate dibasic pH 7.0	20% PEG 3350

Figure 6.2: Sequence alignment of the full-length yeast Tsr1 and the ΔN46Δ410-476 mutant. The sequences are coloured using the Clustal colour scheme.

I scaled up these screen conditions into 24-well sitting drop plates. The crystals obtained were sent to the Diamond Light Source synchrotron and tested for diffraction as before. It was found that crystals in condition 10 gave the best consistent diffraction (on average to about 8 Å resolution). Examples of the crystals can be seen in **Figure 6.3**. Changing the pH of the well solution did not yield better diffracting crystals.

Since changing the crystallisation conditions did not help, multiple replicates of condition 10 were set up and numerous crystals were sent to the Diamond Light

Source (Oxford, UK). The best crystal found diffracted to 3.6 Å. Diffraction data for 270˚ were collected over 1800 images (0.15˚ rotation and 0.1 second exposure) as suggested by EDNA software.

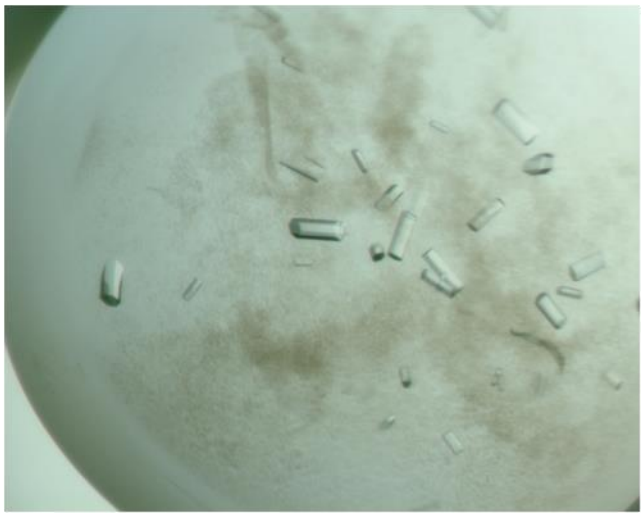


Figure 6.3: Crystals obtained using ScTsr1ΔN46Δ410-476.
These were obtained using the original condition 10. Even, single crystals were observed (about 160 µm in length in the longest direction). They were not improved with additives. Some variation in diffraction was observed when optimising the salt and precipitant concentrations. The best crystal diffracted to 3.6 Å.

EDNA indexes and integrates the diffraction spots observed in the initial 4 images collected in order to obtain the best strategy for full dataset collection. The program ensures that the strategy gives the necessary exposure time and transmission to obtain a sufficient signal-to-noise ratio. It also estimates the completeness and multiplicity predicted to be obtained (Incardona et al, 2009). Thus, it allows collection of the best dataset based on the 4 images used to test diffraction. The collected dataset was further processed as described in the next section.

In order to confirm that the correct protein has crystallised, some of the crystals were collected and dissolved. They were then analysed on a silver-stained SDS-PAGE gel. As shown in **Figure 6.4**, ScTsr1Δ46Δ410-476 was present in the crystals.

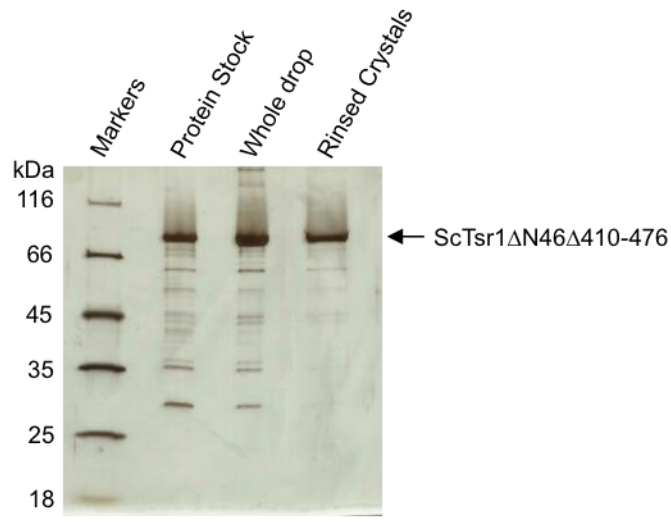


Figure 6.4: Silver-stained SDS-PAGE gel of the ScTsr1ΔN46Δ410-476 protein and the crystals shown in the previous figure.
As controls, the protein stock from the -80˚C freezer and the whole contents of the drop were also analysed. The crystals were rinsed twice in well buffer to remove any protein that was not part of the crystal. They were then dissolved in the SEC buffer and separated on SDS-PAGE. As shown, the crystals contained, almost exclusively, the correct protein. Contaminants and/or proteolytic cleavage products were not seen in the crystals and remained in solution.

6.3 Native Data Processing

The obtained images for the native protein were processed using the XDS software (Kabsch, 2010). In the process, the diffraction spots were indexed and their corrected intensities recorded and integrated. Initially, XDS identified orthorhombic geometry from the diffraction data (space group 16, P222) based on the angles in reciprocal space. This was the highest geometry possible that would still fit the data. However, the processing does not differentiate between the chiral space groups within the Bravais lattice type (i.e. the screw axes are not determined). These were observed in the lattice plane intensity data. Systemic absences (i.e. no intensity) were present in the odd reciprocal planes perpendicular to the lattice vectors in all three directions (Hahn, 1995). Thus, adjacent molecules in the unit cell were translated as well as rotated (a screw axis operation).

Therefore, after scaling the data with SCALA in the CCP4 suite (Winn et al, 2011), Cad was utilised to ensure the correct space group was applied (19, $P2_12_12_1$). The final parameters obtained for this dataset can be seen in **Table 6.3**. Two out of

the three unit cell edges are long, suggesting that the components within are also elongated in one direction. This causes additional difficulties in data collection as the diffraction spots become very close together at certain angles.

Table 6.3: Summary of the parameters obtained after the initial processing of the diffraction data.
The numbers in brackets correspond to the values for the highest resolution shell.

Beamline	I03
Energy (eV)	12028.0
Space group	$P2_12_12_1$
Unit Cell	a = 65.65, b = 173.90, c = 319.75 $\alpha = \beta = \gamma = 90^{\circ}$
Resolution	48.77-3.60 (3.79-3.60) Å
Number of Reflections	437,274
Unique Reflections	43,821
$R_{measured}$ (%)	12.7 (92.2)
Completeness (%)	99.9 (99.5)
I/σI	14.4 (2.9)
Multiplicity	10.0 (9.9)

6.4 Phasing using Molecular Replacement and Heavy Atom Soaking

6.4.1 Molecular replacement (MR) was unsuccessful

As stated in the methods section, solved structures of archaeal eIF5B (PDB ID: 1G7R) and SelB (PDB ID: 4ACA) were used as molecular replacement models in an attempt to solve the ScTsr1ΔN46Δ410-476 structure. These two proteins contain GTPase domains in their N-termini, which are more conserved than other parts of the sequence. A sequence alignment is shown in **Figure 6.5**. The archaeal proteins share 19.0% sequence identity with each other. ScTsr1 shows only 9.1% identity to eIF5B and 11.8% identity with SelB. Normally, a sequence identity of at least 35% is used for MR (Abergel, 2013). However, although the homology is very low in this case, if structural elements look the same in the N-terminal domain, useful phases might be obtained. In the alignment, the N-termini do appear to be most conserved, supporting that Tsr1 contains a GTPase-like domain.

Thus, these structures were used as search models in Phaser (McCoy et al, 2007). Since the GTPase domain is the most likely to be structurally as well as sequentially conserved, different C-terminal truncations of the models were tested in

Phaser as MR search models. In eIF5B the residues following position 443, 328, or 225 were deleted from the PDB file. For SelB, the truncations were at 385, 269, and 178. Also, since the proteins share little sequence identity, the side chains were removed from all the amino acids in the models and replaced with alanines.

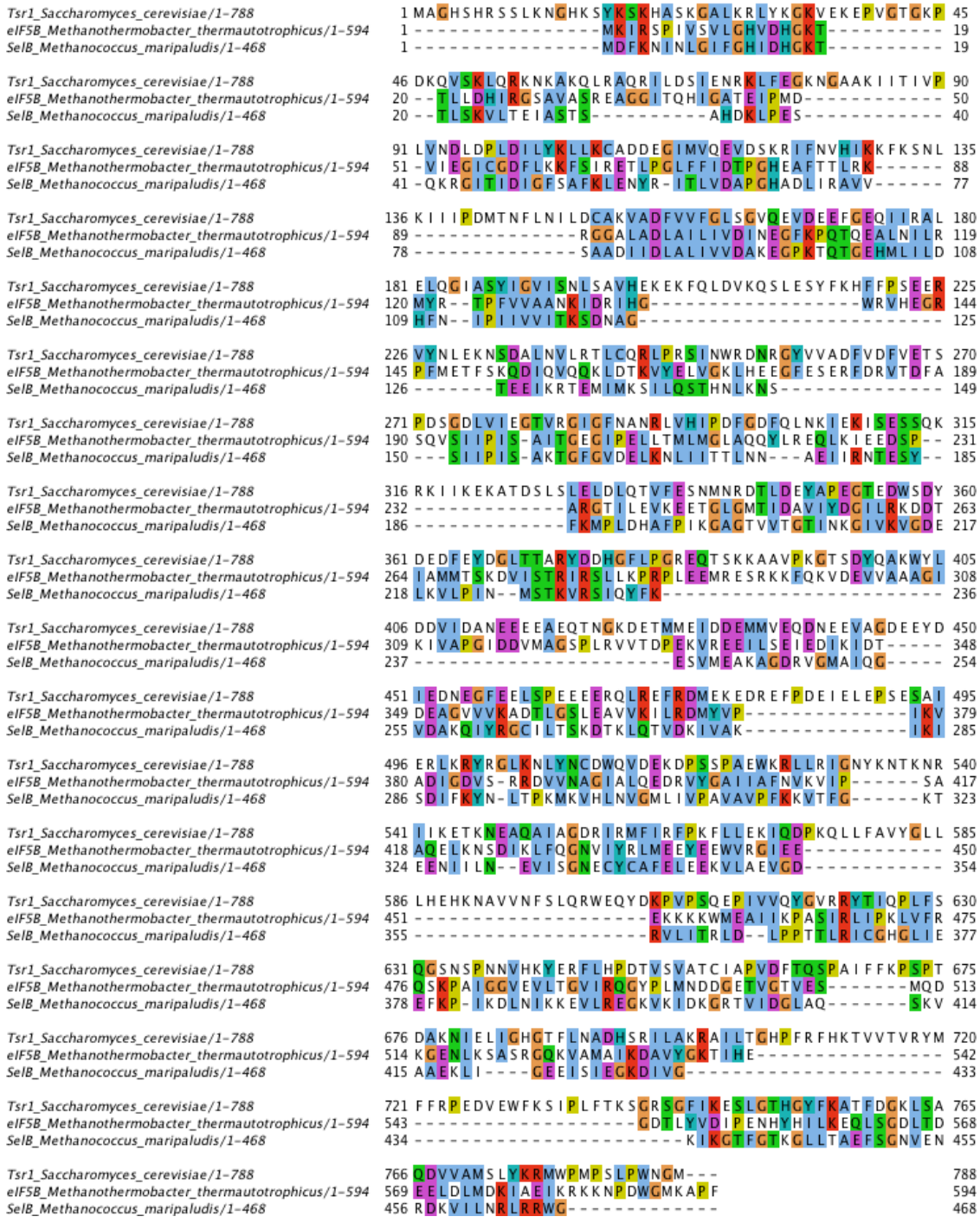


Figure 6.5: Sequence alignment of ScTsr1 and archaeal eIF5B and SelB.
The N-termini are the most conserved. The rest of the protein shows more variability in sequence.

Overall, none of the molecular replacement attempts were successful. The solutions obtained showed lower log-likelihood gain (LLG) and Z-score values than expected for a true solution (Storoni et al, 2004). They were very close to those predicted for a random model. Also, when the maps were examined in Coot (Emsley et al, 2010), no structural elements could be observed in the electron density, which showed small and disjointed segments.

Even though the phasing was unsuccessful, useful information about the crystal was obtained from the preliminary analysis carried out in the CCP4 program suite. Firstly, the Matthews coefficient (V_M) and the likely number of protein molecules in the asymmetric unit (ASU) were calculated using the molecular weight of the protein and the cell dimensions (**Equation 1**).

$$V_M = \frac{\text{Unit Cell Volume}}{\text{Molecular Weight} \times \text{Molecules in ASU} \times \text{Number of ASUs}} \qquad \textbf{Equation 1}$$

The V_M is related to the amount of solvent present in the crystal (**Equation 2**). For most protein crystals it falls between 1.7 and 3.5 $Å^3$/Da with the majority close to 2.15 $Å^3$/Da (Kantardjieff & Rupp, 2003b; Matthews, 1968). Thus, the number of molecules in an ASU that gives V_M closest to 2.15 $Å^3$/Da is usually correct.

Each repeating unit of the crystal (unit cell) is made up of several identical symmetry-related ASUs (which can be made up of several non-identical units). A crystal with the $P2_12_12_1$ space group has to have at least 4 ASUs. Using this number in the V_M equation, the ScTsr1N46Δ410-476 crystal most likely contains 5 protein molecules per ASU, giving a V_M of 2.33 $Å^3$/Da and a solvent content of 47.2% (as calculated using **Equation 2**) (Matthews, 1968). However, as stated above, the V_M can fall anywhere between 1.7 and 3.5 $Å^3$/Da. Thus, although the value using 5 molecules per ASU is most likely, other numbers would also generate a V_M that falls in this range (**Table 6.4**).

$$\text{Solvent Content} = 1 - \frac{1.23}{V_M} \qquad \textbf{Equation 2}$$

Table 6.4: Numbers of molecules per ASU that give V_M values within the accepted range and the solvent content they correspond to.

Molecules per ASU	V_M (Å^3/Da)	Solvent Content (%)
4	2.91	57.8
5	2.33	47.2
6	1.94	36.7

More recently, it has been shown that the resolution of the diffraction pattern correlates with the V_M. Crystals diffracting to a higher resolution have shown lower V_M values (i.e. they have a lower solvent content) (Kantardjieff & Rupp, 2003a). Since the Tsr1 crystals presented here only diffract to 3.6 Å, a larger solvent content could be expected than the one predicted from theoretical values. Thus, 4 molecules per ASU may actually be more likely than the 5 assumed using the equation.

Nevertheless, the unit cell appears to contain 16-20 protein molecules. This would explain the large dimensions observed. Also, Phaser analysed the intensities obtained to determine that there is no twinning evident. If the crystals were twinned, there would be unexpected systemic absences and/or the distribution of the intensities would be different than expected (e.g. fewer very strong/weak and more intermediate intensities). These data can be plotted on a cumulative intensity distribution graph and should follow predicted theoretical curves if no twinning is present. In this case, no twinning was detected. This analysis supports that the quality of the data is good.

6.4.2 Heavy atom soaking did not yield useful derivatives

In order to obtain the phasing information experimentally, some of the crystals were picked and soaked in heavy atom solutions as described in the methods section. Five heavy compounds were used: thimerosal, gold (I) potassium cyanide, lead (II) acetate trihydrate, samarium (III) acetate hydrate, and potassium tetranitroplatinate (II). These were chosen as they display a variety of reactivities.

Samarium and lead form non-covalent bonds with carboxylate groups while the others (mercury, gold, and platinum) form covalent bonds with sulphur atoms and nitrogen atoms in imidazole moieties (Blundell & Johnson, 1976). The ScTsr1ΔN46Δ410-476 construct consists of 686 residues in total. Out of these, 99 are negatively charged (51 aspartates and 48 glutamates) and should interact with samarium and lead compounds. The methionines (9), cysteines (5), and histidines

(14) should bind the other three metals. In addition, they could form other non-specific interactions. Thus, the choice of these five metals provides a variety of potential interactions upon soaking.

The soaked crystals were washed (back-soaked), flash-cooled, and sent to the synchrotron. Prior to data collection, fluorescence scans were performed in order to identify the correct wavelengths to be used for each heavy atom. In this experiment, the crystal is exposed to X-rays of different wavelengths. The heavy atoms are able to absorb the X-rays at certain energies and emit fluorescence based on which heavy atom was used. The energy emission at different wavelengths is measured and the ones necessary to obtain the best anomalous diffraction can be identified. Although theoretical values exist, the scans were performed to ensure that the chemical environment of the heavy atoms was accounted for.

Several datasets were collected for these crystals (**Table 2.10**). However, all of them showed diffraction to 8 Å or worse. The best the dataset (Hg 1) was processed in XDS (in the same manner as the native dataset with the exception that Friedel's law was set to false). This dataset was chosen as it showed diffraction to the highest resolution and the most anomalous signal predicted by Fast_dp automatic processing (Fast_dp performs XDS and SCALA analysis after every data collection using default parameters).

The parameters obtained following analysis are shown in **Table 6.5**. The dimensions and space group remained the same as for the native dataset. The highest resolution used for processing was 8.2 Å. Higher resolution signal was present but significantly lowered the completeness of the data when it was included. Also, the anomalous signal obtained was very low, particularly in higher resolution shells (above 10 Å).

Table 6.5: Summary of the parameters obtained after the initial processing of the mercury soaked crystal.
The numbers in brackets correspond to the values for the highest resolution shell. The anomalous signal represents the mean difference between the structure factor estimates in numbers of standard deviations.

Beamline	I03
Energy (eV)	12298.5
Space group	$P2_12_12_1$
Unit Cell	a = 66.10, b = 174.70, c = 321.70 $\alpha = \beta = \gamma = 90^{\circ}$
Resolution	49.09-8.24 (8.68-8.24) Å
Number of Reflections	20,663
Unique Reflections	3,709
$R_{measured}$ (%)	5.8 (11.6)
Completeness (%)	92.5 (52.3)
I/σI	23.1 (6.3)
Multiplicity	5.6 (2.0)
Anomalous signal	1.05 (0.65)

Nevertheless, the processed and scaled data was still entered into the SHELXC/D/E suite (Sheldrick, 2010) and processed using the default settings. SHELXC was used to analyse the initial data and predict the shifts caused by the heavy atoms. Then, SHELXD builds a substructure based on the locations of the heavy atoms. Finally, SHELXE provides initial phase information based on the substructure and improves them. As expected, not enough anomalous signal was present to obtain the locations of the heavy atoms.

6.5 Purification of Seleno-Methionine labelled Tsr1

The heavy atom soaking was initially unsuccessful. Further trails could be undertaken changing different parameters (e.g. pH, temperature, atoms used). As an alternative approach, purification of seleno-methionine ScTsr1ΔN46Δ410-476 was undertaken. In this case, the heavy atom (selenium) is incorporated directly into the protein. There are 9 methionines in the construct and all of them should be replaced by the heavy derivative. Thus, the positioning of the atom is more specific than soaking and should lead to better crystallisation (if packing is unaffected), as each molecule in the crystal should be more homogeneous. In soaking, the binding of the atoms is often unspecific and can lead to variability.

The protein was expressed in M9 medium supplemented with all the standard amino acid with the exception that methionine was replaced with the seleno-methionine. The expression was performed in B834 (DE3) cells, as they are unable to produce their own methionine and would have to use the stock provided. The two-step purification protocol, developed by Dr. Uma Jayachandran for the native protein, was used. As shown in **Figure 6.6**, pure protein was obtained.

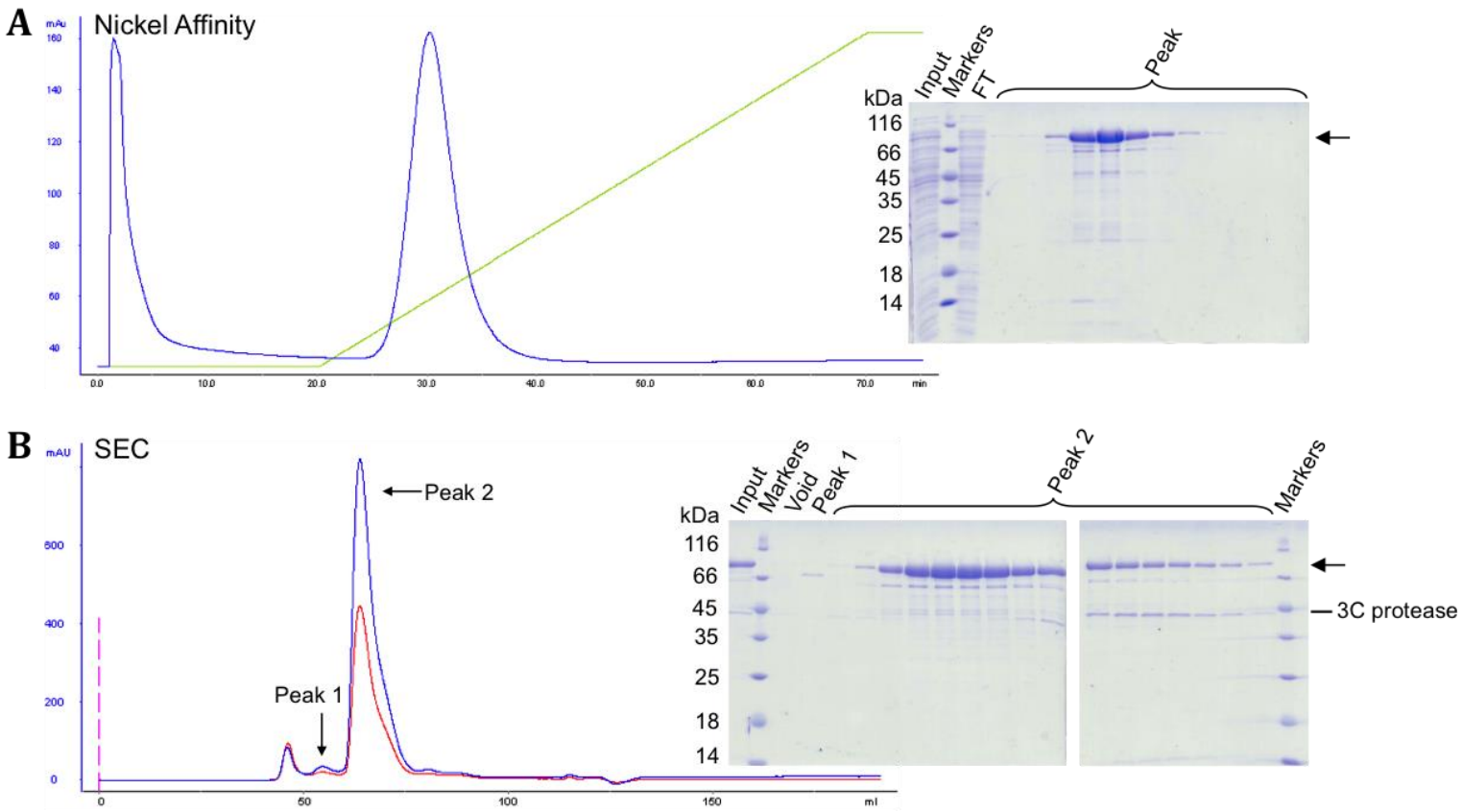


Figure 6.6: Purification of the seleno-methionine labelled ScTsr1ΔN46Δ410-476.

The labels and colours are the same as in previous purification figures.

A) Nickel affinity purification was used as the first step. The protein was eluted with a 10-500 mM gradient of imidazole. The fractions containing the peak were analysed on SDS-PAGE, pooled and dialysed with 3C protease.

B) The protein was then concentrated and separated on SEC. A single peak (2) containing the Tsr1 construct was seen. This overlapped with the 3C protease, as shown in SDS-PAGE. The fractions containing ScTsr1ΔN46Δ410-476 and not 3C protease were pooled, concentrated, and used for crystallisation screens.

6.6 Optimisation of Seleno-Methionine Crystals

Once the heavy atom-labelled protein was purified, it was crystallised in the same conditions (based on condition 10 in **Table 6.2**) as the native protein. The crystals looked similar but less even and the edges were not as sharp (**Figure 6.7**). Multiple crystals were sent to the Diamond Light Source (Oxford, UK) to be tested for diffraction. A fluorescence scan was performed, in the same manner as for the heavy atom soaked crystals, and showed that the selenium was incorporated into the crystal (**Figure 6.8**). This also allowed identification of the best wavelengths for data collection.

Two datasets for the best crystal observed were collected (see **Table 2.11** for strategies): one at the peak wavelength and one at the inflection energy. The datasets were analysed with XDS as for the other crystals. The parameters obtained can be seen in **Table 6.6**.

Table 6.6: Summary of the parameters obtained after the initial processing of the two datasets obtained suing the seleno-methionine labelled protein.
The numbers in brackets correspond to the values for the highest resolution shell. The anomalous signal represents the mean difference between the structure factor estimates in numbers of standard deviations.

Dataset	Peak	Inflection
Beamline	I04	I04
Energy (eV)	12662.0	12660.0
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit Cell	a = 65.75, b = 173.91, c = 319.71 $\alpha = \beta = \gamma = 90^\circ$	a = 66.27, b = 175.14, c = 320.40 $\alpha = \beta = \gamma = 90^\circ$
Resolution	49.02-7.92 (8.35-7.92) Å	49.02-8.96 (9.45-8.96) Å
Number of Reflections	45,158	18,137
Unique Reflections	4,416	3,087
$R_{measured}$ (%)	22.7 (58.2)	21.1 (57.5)
Completeness (%)	98.5 (93.1)	98.1 (92.8)
I/σI	10.4 (3.3)	8.1 (3.5)
Multiplicity	10.2 (6.1)	5.9 (5.5)
Anomalous signal	1.06 (0.78)	0.90 (0.76)

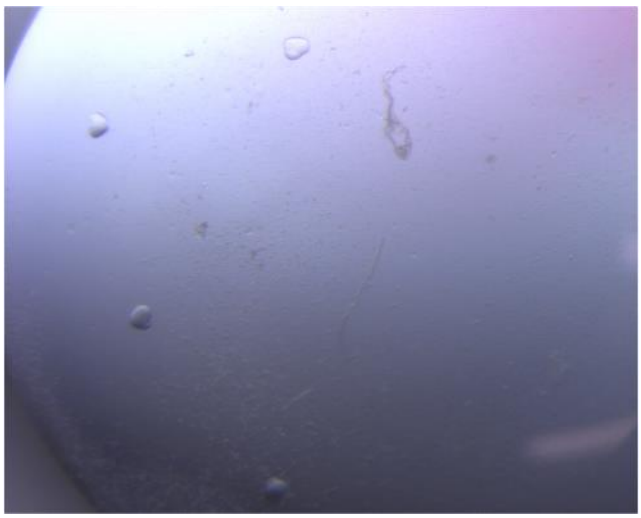
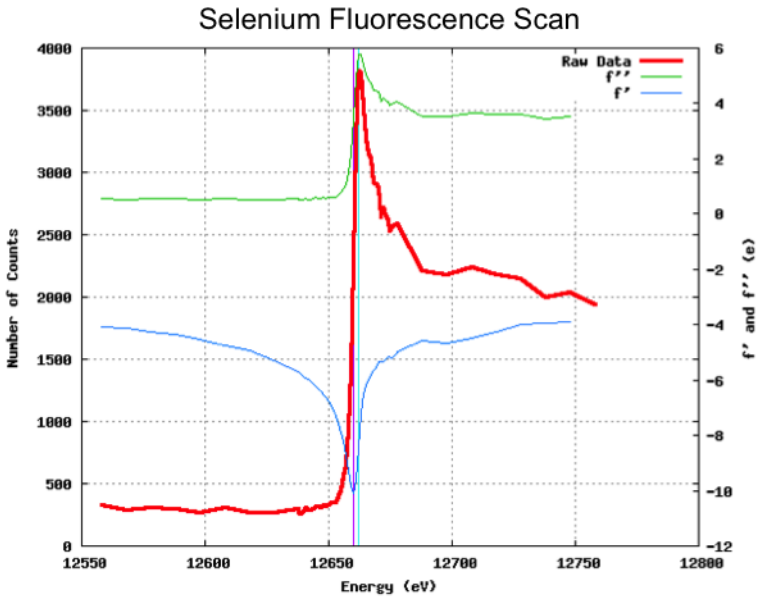


Figure 6.7: Crystals obtained with the seleno-methionine labelled ScTsr1ΔN46Δ410-476.
The crystals are smaller (about 90 µm in length) and the edges are not as sharp. This implies that the crystal is less well ordered when the sulphur atoms are replaced with larger selenium atoms.



	Energy (eV)
Peak	12662.00
Inflection	12660.01

Figure 6.8: Fluorescence scan results for the seleno-methionine labelled ScTsr1ΔN46Δ410-476 crystals.
The experimental data is shown as a red line. The crystal was exposed to X-rays of different energies (x-axis) and the amount of fluorescence released, as the X-rays were absorbed, was recorded (left y-axis). Then, the peak (f′′) and inflection (f′) energies were calculated. They correspond to the energy at which the maximum fluorescence is observed (green) and the point of inflection on the fluorescence curve (blue) respectively. The values obtained from the experiment are listed in the table on the right.

The lower resolution compared to the native dataset suggests that these crystals are less well ordered. Nevertheless, the packing did remain the same as the space group and unit cell dimensions did not change significantly. Thus, the data was entered into SHELXC/D/E suite (Sheldrick, 2010) and processed using the default settings in the same manner as the heavy atom derivatives. Unfortunately, the resolution was too low to obtain phase information.

Thus, optimisation of the labelled crystals was undertaken. The protein was re-screened using the Hampton Index and Magic I (Munich crystal platform) screens. The two initial hits are shown in **Table 6.7**. The number of hits significantly decreased from the native protein supporting that the crystal formation might be affected by the modification.

Table 6.7:The hits obtained using seleno-methionine-labelled ScTsr1ΔN46Δ410-476

Condition	Buffer	Salt	Precipitant
1	50 mM MES pH 6.0	0.2 M Ammonium acetate	4% MPD
2	50 mM MES pH 6.0	0.2 M Sodium malonate	18% PEG 8000

The two conditions were taken forward into 24-well sitting drop plates. Some crystals have been obtained. They were very small but some of them did diffract. Further optimisation is needed to increase their size.

6.7 Discussion

6.7.1 Crystallographic data collection for native crystals was successful

The original loop-out construct gave multiple hits in commercial screens. However, these did not diffract to a high enough resolution for structure solution. Removing the N-terminal portion of the protein aided in the formation of better crystals. This could be immediately seen in the commercial screens where more hits were obtained. These results support the prediction that this portion of the protein is unstructured.

The best native data obtained from the crystals showed diffraction up to 3.6 Å resolution. Data processing showed high completeness (99.6%) and a high signal-to-noise ratio (I/σI), even in the highest resolution shell (I/σI = 2.97). The data also

displayed a relatively low R_{meas} value, which suggests that multiple observations of the same data converge well.

Further data processing would require a solution to the phase problem. **Equation 3** shows the Fourier transform that needs to be applied to every point in order to obtain the electron density (ρ). The F($h\ k\ l$) values can be calculated from the spot intensities (I($h\ k\ l$) in **Equation 4**). However, the phases (α) cannot be calculated from the diffraction data and need to be known prior to data analysis. Thus, they must either be based on similar solved structures or experimentally determined.

$$\rho(x,y,z) = \frac{1}{V}\sum_h \sum_k \sum_l |F(h\ k\ l)| \exp[-2\pi i(hx + ky + lz) + i\alpha(h\ k\ l)] \quad \textbf{Equation 3}$$

$$I(h\ k\ l) \propto |F(h\ k\ l)|^2 \quad \textbf{Equation 4}$$

6.7.2 Phasing information has not yet been obtained

Initially, to obtain phasing information, molecular replacement was attempted. For this purpose, the solved archaeal structures of eIF5B and SelB were used. The eIF5B GTP-binding protein (also known as Fun12) is an initiation factor, which aids in the ribosome subunit joining (Kolakofsky et al, 1968) and is involved in the translation-like cycle that allows D site cleavage to occur. Its interaction with the ribosome is blocked by Tsr1 (Strunk et al, 2011). SelB is a translation elongation factor that incorporates seleno-cysteines into forming protein chains (Leibundgut et al, 2005). Both proteins display ribosome binding sites in a similar location to that of Tsr1 and, thus, might have some structural similarities.

The two proteins do not share high sequence homology with Tsr1 (about 10% sequence identity). They are more similar to each other (19% sequence identity and RMSD of 14.6 Å between archaeal eIF5B and SelB structures) and display a similar fold, particularly in the N-terminal GTPase domain (**Figure 6.9A**). The amino acid side chains were removed for the modelling and three truncations keeping the GTPase domain were created (eIF5B at 443, 328, or 225 and SelB at 385, 269, and 178) to decrease the potential differences in fold.

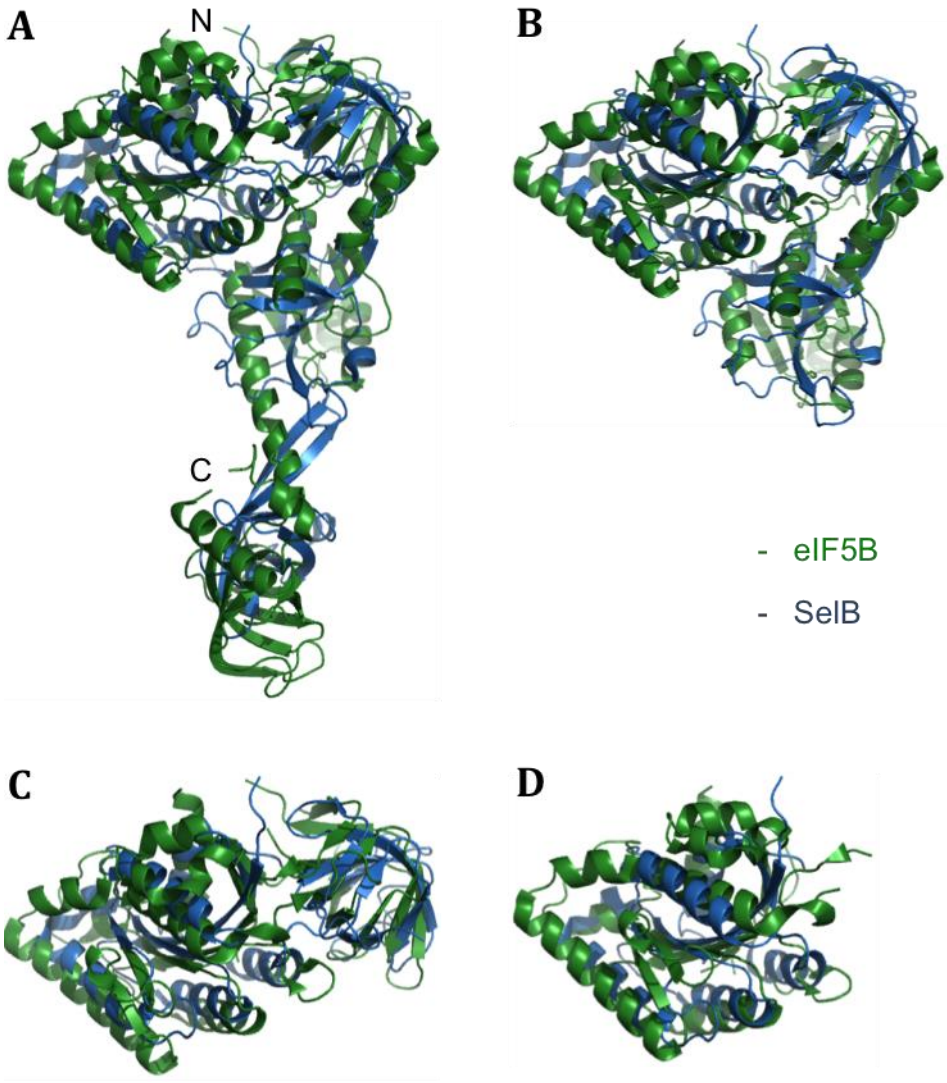


Figure 6.9: Overlay of the full-length and truncated archaeal eIF5B (PDB: 1G7R) and SelB (PDB: 4ACA).

The proteins are shown in green and blue respectively.

A) Overlay of the full-length proteins. The N-terminal domain appears to be more similar in structure. The structures shown an RMSD of 14.6 Å.

B) Overlay of the first truncation (at 443 and 385 respectively). The RMSD is decreased to 12.7 Å by removing the C-terminal domain.

C) Overlay of the second truncation (at 328 and 269 respectively). The RMSD is further decreased to 8.3 Å.

D) Overlay of the last truncation (225 and 176 respectively). The final RMSD is 5.9 Å

None of the molecular replacement trials yielded any usable maps. This could be due to the low sequence identity between the proteins. Usually, molecular replacement utilises models with at least 35% sequence identity. This translates to an average RMSD of 1.5 Å in the main chain carbon atoms (Abergel, 2013). Sequence divergence highly affects the structure similarity. For example, the archaeal eIF5B

and SelB structures show 19% sequence identity and an RMSD of 14.6 Å. This is decreased to 12.7, 8.3, and 5.9 Å aligning each of the truncations respectively (**Figure 6.9B-D**), showing that the more conserved regions display higher structural similarities. As Tsr1 only shares about 10% sequence identity with these proteins, finding a solution was unlikely.

Nevertheless, newer structures are being published. Recently, structures of the eukaryotic *C. thermophilum* eIF5B protein in different nucleotide-bound states with a variety of conformations have been released (Kuhle & Ficner, 2014). This protein shares only about 13% sequence identity with Tsr1 and is unlikely to give a solution. However, it is a longer protein than the archaeal variant (over 1092 residues vs. 594) and could be more similar in fold to Tsr1. Due to time constraints, molecular replacement using these structures has not been attempted but should be performed in the future.

Subsequently, heavy atom soaking was attempted. These atoms, like all others in the crystal, would contribute to the diffraction pattern. The contributions would be relatively high as heavy atoms are electron dense and have a higher scattering power compared to atoms found in polypeptide chains. The changes in spot intensities from the native data would allow preliminary phase information to be estimated. However, for this method to work, heavy atoms must be incorporated. They would have to enter the crystal lattice via the solvent channels and occupy uniform spaces in each unit cell. This can prove difficult as the binding is often unspecific and depends largely on the soaking conditions.

Different heavy atom derivatives were used to soak the native crystals (as described in the methods section). The crystals were placed in the heavy atom solutions and then back-soaked in cryo-protectant. The fluorescence scans showed that the atoms were present in the crystal and the optimal wavelengths for data collection were identified. Nevertheless, the anomalous signal obtained was very low and the maximum resolution was decreased. This suggests that the crystals showed more disorder and few atoms were incorporated. The higher disorder was likely caused by heterogeneous occupancy.

Even though useful derivatives were not obtained, this method can be further optimised. Relatively few heavy atom salts were screened. Other compounds could

be used. Also, the binding of the atoms is highly affected by the reaction dynamics. Thus, different buffer pH and temperatures can have large effects (Blundell & Johnson, 1976). These could be further screened in the future.

In the meantime, protein containing the heavy atoms within its sequence was purified (seleno-methionine labelled). The construct behaved in the same manner as the native protein in purification. It also crystallised in the same conditions. However, the crystals did not diffract to the same resolution (highest resolution was about 8 Å). This could have been caused by slight differences in the crystal order due to the larger atom (selenium) replacing the sulphur.

Thus, multiple rounds of crystal optimisation were undertaken. Due to time constraints, many of these have not been fully optimised. The crystals are still too small to give useful diffraction. However, this is a promising path and will likely yield results. The advantage of using seleno-methionine labelled protein is that the heavy atom is incorporated into the sequence. Therefore, its positioning should be uniform within each unit cell.

6.8 Concluding remarks

Overall, preliminary steps have been taken to solve a eukaryotic Tsr1 structure. Good native crystallographic data are available. There is a good starting point for obtaining experimental phases as a seleno-methionine derivative has been created and is in the process of being crystallised. Also, many new protein structures are being published and some of them might prove to be good molecular replacement search models in the future. Once the structure is obtained, insight into the Tsr1 function in ribosome biogenesis will likely follow.

6.9 Update

Since the initial submission of this thesis in October of 2014, further heavy atom soaks were performed. Crystals containing tantalum bromide clusters were obtained and initial phases were calculated. The methods and results for this

additional work can be seen in **Appendix V** while the resulting electron density cloud is shown below in **Figure 6.10**.

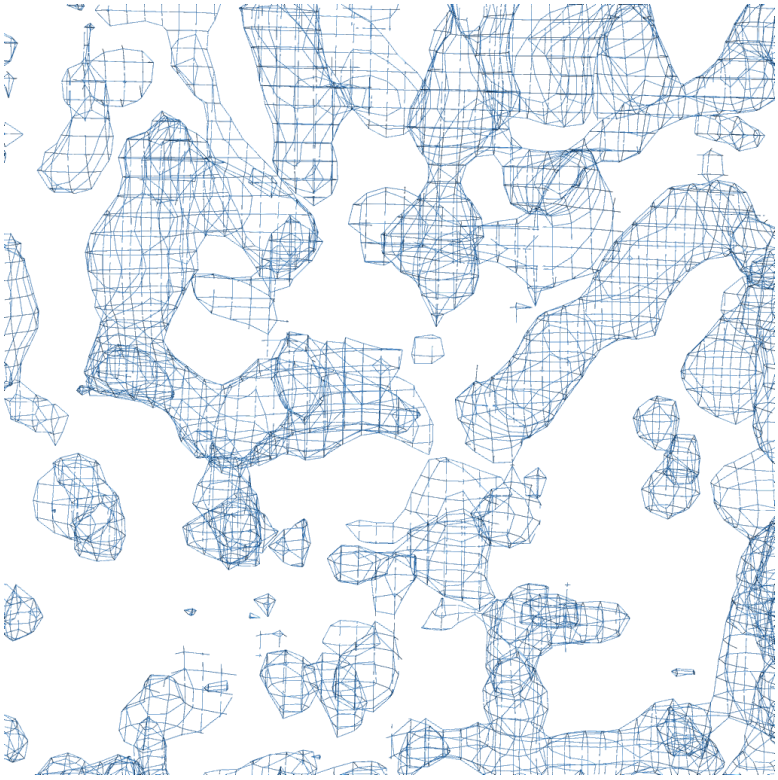


Figure 6.10: A sample image of the electron density obtained using multi-wavelength anomalous dispersion (MAD) experimental phasing.
This is a 2Fo-Fc map where the σ value was set to 1.0. The map shows good connectivity and empty space for the solvent channels. The resolution of the derivative crystal (4.2 Å) prevents easy model building into the density. In the future, model building and phase extension will be used to improve the quality of the map.

7 CONCLUSIONS AND FUTURE PROSPECTS

7.1 Recombinant Ribosome Biogenesis Factors were Successfully Purified and Characterised

Over the course of the project, three eukaryotic ribosome biogenesis factors have been successfully expressed and purified from *E. coli* strains. The protocols were optimised and found to be reproducible. Nob1 (from *S. cerevisiae* and *C. thermophilum*) and human Dim2 were found to be stable using a thermal denaturation assay. Different constructs for each protein were also obtained. The protocol for the Tsr1 constructs was optimised by Dr. Uma Jayachandran, who showed that these were also stable using thermal denaturation.

The majority of the experiments performed on Nob1 used the yeast full-length construct. The recombinant protein was found to have its RNA binding activity and to be active in an *in vitro* D site cleavage assay. However, the RNA binding predicted to occur between Nob1 and helix 40 (Granneman et al, 2010) was not observed. It is possible that this interaction is mediated by other parts of the pre-ribosome. Otherwise, the conformations of the artificial rRNA fragments might be preventing the interaction from occurring *in vitro*. Thus, although the zinc ribbon domain of the archaeal protein was shown to bind to helix 40 (Veith et al, 2012), it remains unclear if this is the case in eukaryotes.

Instead, the yeast protein was found to interact with a sequence (AGGA) immediately upstream of the D site and with one about 100 bases downstream (the ITS1 sequence). Previously, rRNA fragments encompassing these sites have been identified (Lamanna & Karbstein, 2009) but the specificity was not determined to this level of resolution. The two modes of binding appeared distinct in ITC and are likely performed independently by the two separate domains. It is possible that the PIN domain binds the D site while the ITS1 sequence interacts with the zinc ribbon.

Further experiments need to be performed in order to determine if both sequences can interact with the protein at the same time. For example, a competition assay could be performed in ITC. For this purpose, the protein would firstly be

saturated with one of the RNAs. Then, the second one would be injected. Its affinity should not be changed by the presence of the other RNA if they bind independently. If they bind co-operatively, the affinity would be higher and if there were competition for the interaction, the affinity would be lower. During these experiments, it is important to keep in mind that in archaea, the zinc ribbon domain interacts with helix 40. This would add more complexity as it is unclear if and how all three interactions would co-exist.

Dim2 constructs have also been successfully purified. The recombinant human and truncated yeast Dim2 proteins appear to be able to interact with RNAs containing the predicted binding site (GGAUCA). Both Dim2 constructs were found to precipitate upon addition of D site RNA. The precipitation suggests a binding event. It was only observed when the RNA was single-stranded supporting that a specific interaction is occurring. This also supports the hypothesis that a hairpin is not formed at this site during ribosome biogenesis (Lamanna & Karbstein, 2009). The interaction was disrupted at higher pH and in higher salt. In the future, the binding experiments should be performed with shorter RNA sequences in order to prevent precipitation. Also, the sequence requirements for binding should be tested by introducing point mutations in the same manner as was done with Nob1.

The yeast truncated Dim2 protein is also able to interact with Nob1. These results support that the recombinant proteins are correctly folded and can perform their predicted functions. The full-length yeast protein could not be purified as it was heavily contaminated with bacterial rRNA and ribosomal proteins in all protocols attempted.

Finally, *S. cerevisiae* Tsr1 constructs were designed and purified by Dr. Uma Jayachandran. However, the function of this protein remains unknown and a specific recognition site has not been mapped. Cross-linking has shown that it is in close proximity to long and disparate sequences of the 20S pre-rRNA, including proximal sequences to the D site (Granneman et al, 2010). Thus, it is unknown what short sequence could be designed for testing of Tsr1 folding. In the future, EMSAs could be performed with different parts of the identified sequences in order to pinpoint the most important fragment.

7.2 Insight into the 20S pre-rRNA Processing has been Obtained

Identification of the Nob1 RNA recognition sequences provides insight into the last step of maturation of the small ribosomal subunit. The two distinct binding sites (D site and ITS1) with different binding modes suggest a specific order of events and interactions that allow the nuclease domain to be positioned correctly.

Also, the site identified upstream of the cleavage event overlaps the sequence predicted for Dim2 binding. Thus, the positioning of the nuclease domain is dependent on the other factors in the processing pathway. It is likely that Dim2 needs to be released before the Nob1 PIN domain can interact with its target.

Finally, the yeast Nob1 protein is able to bind to the D site cleavage product in the same manner as to the substrate but with a slightly higher affinity. This suggests that additional factors are required for the release of the mature particle. This may form another checkpoint for the fidelity of the cleavage and a further level of control as Nob1 cannot bind to another site and cleave it until it is safely released from its previous target.

All of these data could be added to the proposed translation-like cycle (**Figure 1.4** in Chapter 1), which shows that the pre-40S subunit has to interact with the 60S subunit (mediated by Fun12 GTPase) in order for the small subunit to undergo final maturation (Lebaron et al, 2012; Strunk et al, 2012). **Figure 7.1** summarises the steps thought to occur based on the data presented in this thesis.

The data suggest that Nob1 binds to an ITS1 sequence while Dim2 interacts with the D site. The two proteins also interact with each other at this time. Then, when the rRNA is determined to be ready for final maturation (possibly through the translation-like cycle), Dim2 has to be released by an unknown mechanism. The timing of this release with respect to Fun12 and 60S subunit binding is also unknown. It is possible that GTP hydrolysis performed by Fun12 and the conformational changes associated with it could cause Dim2 to be released. After Dim2 is removed, the Nob1 nuclease domain would be able to interact with the D site and to cleave it. Then, all the factors, including Nob1, would be released, again through an unknown mechanism.

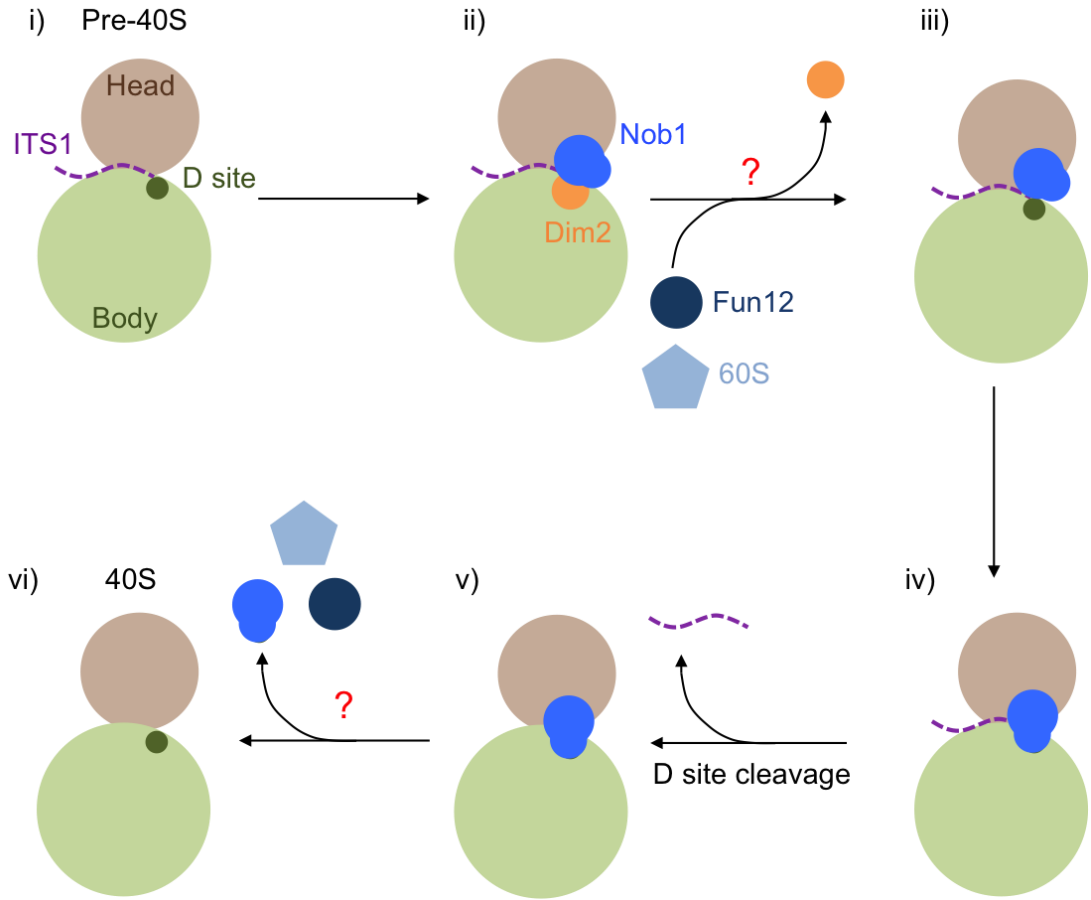


Figure 7.1: Model of the D site cleavage combining data obtained in this project and the translation-like cycle model previously described (see Figure 1.4).
The colouring is consistent with the model shown in **Figure 1.4**. Only the pre-40S subunit is shown and it is divided into the head (brown) and body (green) domains. The D site is shown in dark green and the ITS1 RNA is shown as a dashed purple line. Nob1 is shown in blue and Dim2 is orange. Fun12 (dark blue circle) and the 60S RNA (light blue pentagon) are also indicated. The pre-40S subunit (i) is bound by Dim2 near the D site and by Nob1 in the ITS1 (ii). Then, Dim2 is released by an unknown mechanism (iii). The Fun12 GTPase and 60S subunit would have to bind to allow the translation-like cycle to occur (Lebaron et al, 2012; Strunk et al, 2012). It is unclear if this would happen before or after Dim2 release. Nob1 can now bind to the D site with its nuclease domain (iv) and perform the cleavage (v). Finally, the Rli1 ATPase separates the ribosomal subunits (Strunk et al, 2012) and the accessory factors (including Nob1) are released through an unknown mechanism (vi).

7.3 Crystallisation of the Factors has been Successful

Loop-out constructs, which removed flexible loops joining two folded domains, were created for two proteins (Nob1 and Tsr1) in order to increase the chances of obtaining crystals. This method has been used in the past (Derewenda, 2004) and could prove to be useful for crystallisation of other ribosome biogenesis

factors whose structures have not been solved to date. During the course of the project, it did allow the yeast Tsr1 to be successfully crystallised. A good dataset has been obtained and processed. Although the structure of Tsr1 has not yet been solved, the seleno-methionine construct has been purified and heavy atom derivatives have been obtained. Structural information would aid in distinguishing the function of this protein.

Future prospects for further Nob1 crystallisation trials have also been established. These include co-crystallisation with RNA and/or Dim2 using the loop-out constructs and different RNA sequences identified in ITC.

Once a eukaryotic Nob1 structure is obtained, it can be used to further understand the D site cleavage. It can also be used to identify potential small molecule inhibitors of its function. Cell proliferation in different types of cancer (e.g. ovarian, liver, and lung cancers) has been shown to be decreased when Nob1 is knocked down (Li et al, 2014; Lin et al, 2011; Lu et al, 2011). Thus, drug molecules that target Nob1 function could provide a new cancer treatment with broad specificity. Obtaining a eukaryotic Nob1 structure would be the first step for achieving this.

7.4 A System for Observing D site Cleavage *In Vitro* has been Optimised

Previously, an *in vitro* D site cleavage assay using a long RNA substrate was published (Lebaron et al, 2012; Pertschy et al, 2009). Here, this assay was optimised and used on shorter substrates successfully. As with the published cleavage, the activity of Nob1 was shown to be low in the simplified environment. However, the system can be used successfully with recombinant proteins to test for aiding or inhibition of the D site cleavage.

Nob1 and Dim2 were found to interact with overlapping D site RNA sequences. However, the sequence requirements for Dim2 were not tested (due to the precipitation issue) and the actual positioning is predicted based on the archaeal Dim2 structure in complex with bacterial rRNA (Jia et al, 2010). Nevertheless, it is likely that the two proteins compete for binding.

If a condition giving soluble Dim2-RNA complexes could be found, the competition for binding with Nob1 could be observed in ITC (in the same manner as described for the RNA competition previously). Then, the D site RNA bound to Dim2 could be used in the cleavage assay. It has been shown that Dim2 helps Nob1 to make rRNA contacts (Woolls et al, 2011). However, this might not be true at the D site, as the binding sequences appear to overlap. If Nob1 cleavage becomes more efficient in the *in vitro* system when Dim2 is included, its role in aiding the cleavage would be clarified. Otherwise, it could be preventing the cleavage from happening until the right moment.

Also, different types of substrates could be tested. It is possible that the RNA sequence and secondary structure have an effect of the efficiency of the Nob1 cleavage. Thus, different RNAs should be initially probed by primer extension to obtain secondary structure information. Then, they could be used in the cleavage assay to test for changes in Nob1 activity. Other sequences interacting with the zinc ribbon domain, like ITS1, could also affect the PIN domain nuclease function and should be included in some experiments. The identified RNA mutation that prevents Nob1 binding (AGGA to AAAA) should also be incorporated into some of the substrates to see if the interaction is necessary for cleavage.

Finally, in the future, this assay could also be used to verify any potential Nob1 inhibitors identified. The ability of the small molecules to prevent cleavage would need to be tested *in vitro,* as this would provide a very simple system that would be easier to analyse than an *in vivo* system where other factors might obscure the relative activity of the inhibitors.

7.5 Concluding Remarks

Overall, this project was successful in obtaining insight into the late stages of small subunit pre-rRNA maturation. The previously identified interaction between Nob1 and Dim2 was confirmed using recombinant proteins. However, it was also shown that the two proteins might compete for the D site. Recent cross-linking data in growing cells performed by the Tollervey lab (personal communication with Prof. Tollervey) suggest that this competition also exists *in vivo*. The experiments

presented in this thesis identified conditions that affect the interaction between Dim2 and Nob1 or D site rRNA fragments (high salt and pH). This could be useful in the future for the formation of stable complexes.

The initial Nob1 crystals have prospects for optimisation. The structure would provide a good starting point for the identification of inhibitors. Then, the cleavage assay described in this project could be used to test the effects of these inhibitors. The assay could also be expanded to study the effect of Dim2 on Nob1 nuclease activity.

Although the function of Tsr1 is unknown, large progress has been made in obtaining structural data. A native dataset has been obtained and steps have been taken to solve the phase problem. Overall, it is likely that the structure will be solved in the near future providing further insight.

8 ACKNOWLEDGEMENTS

I am deeply indebted to my main supervisor, Dr. Atlanta Cook, for providing such an interesting project and for all of her help and advice throughout the years. Further appreciation is extended to my second supervisor, Prof. David Tollervey, for sharing in ideas and for his thoughtful comments on the project, and to the rest of my PhD committee (Dr. Bettina Boettcher and Prof. Paul Barlow). I would also like to extend thanks to the University of Edinburgh for giving me this opportunity and the Wellcome Trust for funding.

All members of the Cook lab have helped me a lot, but special appreciation is extended to Dr. Uma Jayachandran who performed most of the Tsr1 cloning and purifications. Additionally, to the former member, Anees Basha, who purified the yeast Dim1 construct. I would also like to thank our summer intern, Michal Merdas, who aided in the purification and crystallisation screening of the Nob1 mutant proteins.

Special thanks are extended to Dr. Nick Watkins at the University of Newcastle for sending the original human protein construct and Dr. Andrew Leech at the University of York for performing the SEC-MALS experiments and subsequent data analysis. Also, to Dr. Sander Granneman at the University of Edinburgh who kindly shared his Fap7 construct and to his colleague, Dr. Ralph Hector, who provided technical support with the primer extension assay.

I would also like to acknowledge all of the people who assisted in data analysis and day-to-day chores during the duration of my PhD. In particular, all of the EPPF, BCF, and Swann floor 3 staff were instrumental throughout the years. Finally, everyone who attended the lab meetings and provided feedback on my presentations is also dearly thanked.

9 APPENDICES

9.1 Appendix I: Primers

Primer Sequences (5′-3′)

ScNob1 gene amplification for conventional cloning
(Restriction sites are shown in green and extra two bases are shown in blue)

ScNob1 forward primer:
CCCCCATGGGCACCGAAAACC

ScNob1 reverse primer:
CCCCCTCGAGACTTCTCCTTT

ScDim2 gene amplification for LIC for 3C vectors
(Overlap with plasmid is shown in red and the extra three bases are shown in blue)

ScDim2 forward primer:
CCAGGGGCCCGACTCGATGGTTGCGCCTACTGCTTTG

ScDim2ΔN51 forward primer:
CCAGGGGCCCGACTCGGCAAAAGAAGAGGTAGAGG

ScDim2ΔN61 forward primer:
CCAGGGGCCCGACTCGAGCCGGAAAACGCACG

ScDim2ΔN91 forward primer:
CCAGGGGCCCGACTCGAAAATAAAATTTGAATCGAGAAAGATCATGGTTC

ScDim2ΔN99 forward primer:
CCAGGGGCCCGACTCGATGGTTCCACCACACAGAATG

ScDim2 reverse primer:
CAGACCGCCACCGACTGCTTAGTAGCGTTCTTTTAATCTAGATGCAAC

ScDim2 gene amplification for LIC for TEV vectors

ScDim2 forward primer:
CCAGGGAGCAGCCTCGATGGTTGCGCCTACTGCTTTGAAAAAGGCTAC TGTG

ScDim2 reverse primer:

GCAAAGCACCGGCCTCGTTAGTAGCGTTCTTTTAATCTAGATGCAA
CGGTACGTAAGTTCC

pUC19 amplification

Forward:
GAATTCACTGGCCGTCGTTTTACAACGTCG

Reverse:
GCTATGACCATGATTACGCCAAGCTT

Human proteins into 3C vectors (LIC)
HsNob1 (WT and D10N)
Forward:
CCAGGGGCCCGACTCGATGGCTCCAGTGGAGCACGTTGTGGC

Reverse:
CAGACCGCCACCGACTGCTTACCTTTTCTTCACAAACTTCTTTCTG
GAAGCGTTGGGATTTAAG

HsDim2ΔN32

Forward:
CCAGGGGCCCGACTCGGCTGAACAGCTGTCCGCAGCAGGAG

Reverse:
CAGACCGCCACCGACTGCTTAGAATCGATCTGCTGATCTGCTAGCC
ACAGC

HsDim2ΔN43

Forward:
CCAGGGGCCCGACTCGGATGCGGGCCGCATGGACACAGAG

Reverse:
CAGACCGCCACCGACTGCTTAGAATCGATCTGCTGATCTGCTAGCC
ACAGC

pET28α amplification

Forward:
CATGGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGG
GGAATTGTTATC

Reverse:
CTCGAGCACCACCACCACCACCAC

HsNob1 (WT and D10N) into pET28α (G-block cloning)

Forward:
TTTAACTTTAAGAAGGAGATATACCATGGCTCCAGTGGAGCACGT
TGTGGC

Reverse:
TGGTGGTGGTGGTGCTCGAGCCTTTTCTTCACAAACTTCTTTCTGG
AAGCGTTGGGATTTAAG

Primer sequence for primer extension experiments (5′-3′)

AGCTTGCATGCCTGC

RT-PCR Primers

Forward:
CGGAGTGTGGCTCAGCCC

Reverse (77 bases):
TGGTCGGAGTGGCAGGATTC

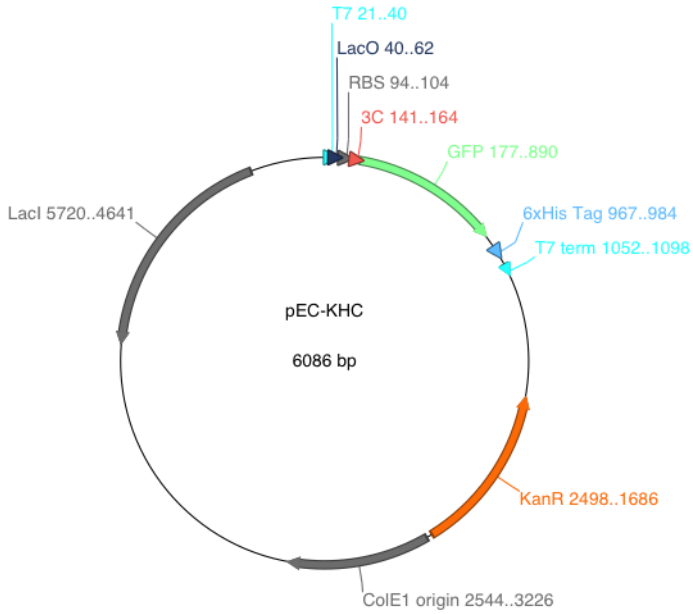
Reverse (99 bases):
CCCAAGGGGTTGCCGACAATC

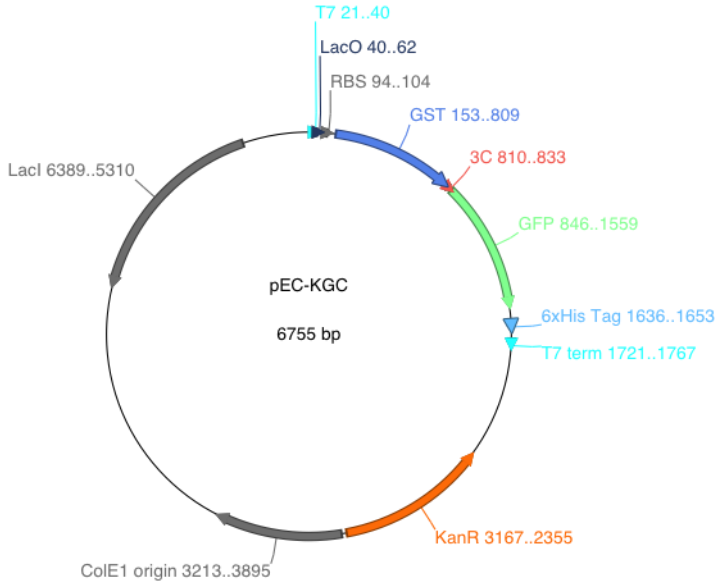
Reverse (109 bases):
GTTTAGAGGCCCCAAGGGGTT

Reverse (121 bases):
CCCTCAAGACCCGTTTAGAGGC

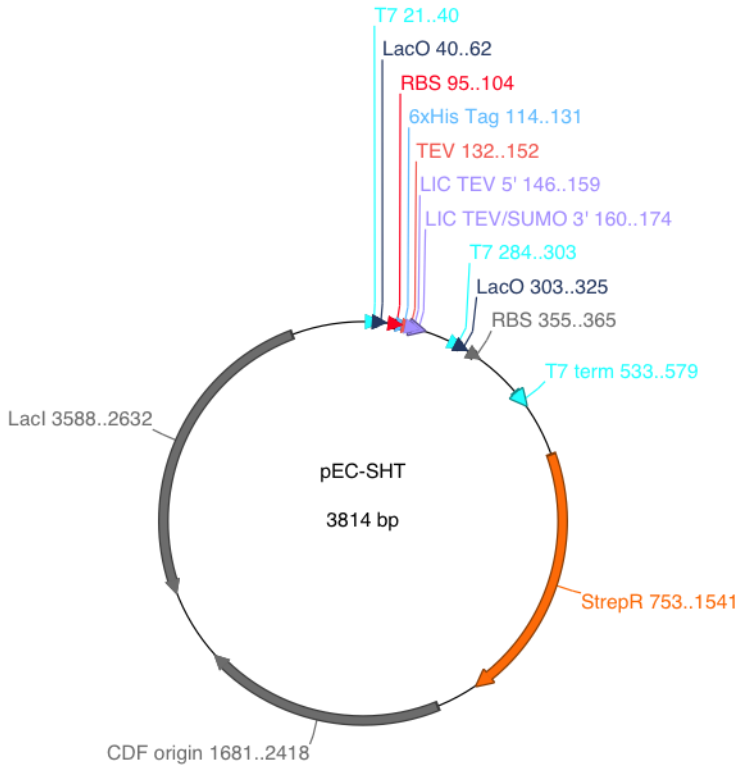
9.2 Appendix II: LIC Vectors

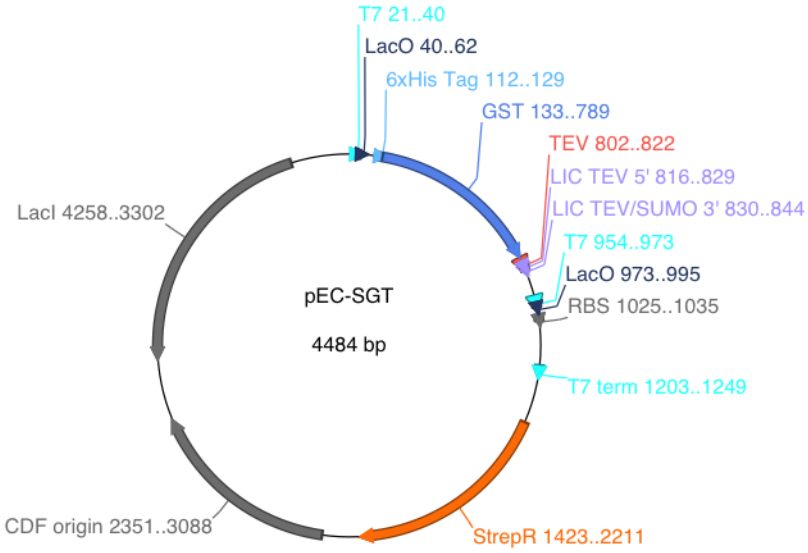
The pEC-KHC and pEC-KGC plasmids have kanamycin resistance (K) and a 3C protease site (C) incorporated as shown below:





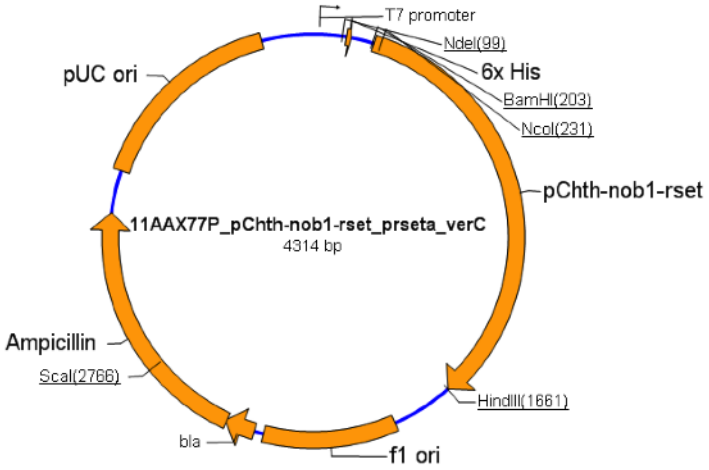
The pEC-SHT and pEC-SGT plasmids have streptomycin resistance (S) and a TEV protease cleavage site (T) as shown on the next page:

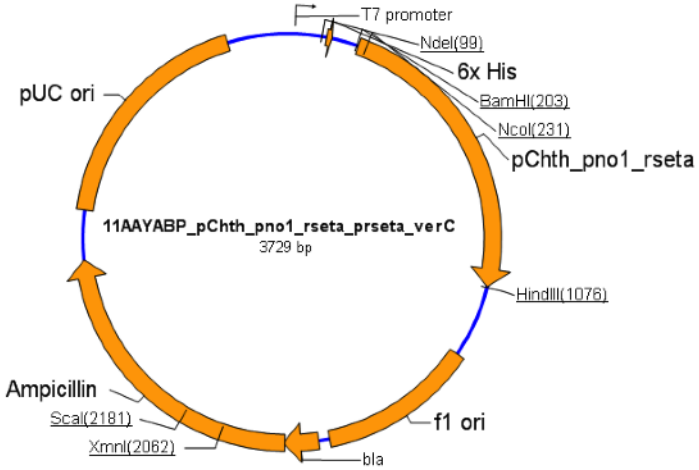

T7 21..40
LacO 40..62
RBS 95..104
6xHis Tag 114..131
TEV 132..152
LIC TEV 5' 146..159
LIC TEV/SUMO 3' 160..174
T7 284..303
LacO 303..325
RBS 355..365
T7 term 533..579
LacI 3588..2632
pEC-SHT
3814 bp
StrepR 753..1541
CDF origin 1681..2418


T7 21..40
LacO 40..62
6xHis Tag 112..129
GST 133..789
TEV 802..822
LIC TEV 5' 816..829
LIC TEV/SUMO 3' 830..844
T7 954..973
LacO 973..995
RBS 1025..1035
LacI 4258..3302
pEC-SGT
4484 bp
T7 term 1203..1249
CDF origin 2351..3088
StrepR 1423..2211

9.3 Appendix III: Additional Plasmids

The *C. thermophilum* proteins were ordered as codon optimised plasmids from GeneArt. The plasmid diagrams for CtNob1 and CtDim2 (or Pno1) are shown below:





9.4 Appendix IV: Transcription Sequences

***In vitro* transcription sequences ordered**
(blue – T7 promoter, red – extra bases)

Homo sapiens D site
Top: TAATACGACTCACTATAGGGAAGGATCATTAACGGAGCCCG
Bottom: CGGGCTCCGTTAATGATCCTTCCCTATAGTGAGTCGTATTA
Saccharomyces cerevisiae D site
Top: TAATACGACTCACTATAGGGAAGGATCATTAAAGAAATTTA
Bottom: TAAATTTCTTTAATGATCCTTCCCTATAGTGAGTCGTATTA

***In vitro* transcription sequences cloned**

Saccharomyces cerevisiae H40 (short)
TAATACGACTCACTATAGGGATCGGTTTCAAGCCGATCCC

Saccharomyces cerevisiae H40 (intermediate)
TAATACGACTCACTATAGGGCCTTAACCTACTAAATAGTGGTGCTA
GCATTTGCTGGTTATCCACTTCTTAGAGGGACTATCGGTTTCAAGC
CGATGGAAGTTTGAGGCCC

Saccharomyces cerevisiae H40 (long)
TAATACGACTCACTATAGGGTTCTTAGTTGGTGGAGTGATTTGTCT
GCTTAATTGCGATAACGAACGAGACCTTAACCTACTAAATAGTGG
TGCTAGCATTTGCTGGTTATCCACTTCTTAGAGGGACTATCGGTTT
CAAGCCGATGGAAGTTTGAGGCAATAACTCCC

Saccharomyces cerevisiae D site
TAATACGACTCACTATAGGGATCATTAAAGAAATTTAATAATCCC

9.5 Appendix V: Tsr1 MAD Phasing

Materials and Methods

Native crystals were soaked in 1 or 2 mM tantalum bromide for 0.5, 1, 3, 6, or 16 hours. This was performed in the same manner as previously described for other metals. The back-soaking step was also used. Data was collected remotely on the I02 or I03 beamline at the Diamond Light Source synchrotron in Oxford. As with the other metals, a fluorescence scan was performed and data collection was performed at the peak, inflection, or high remote energy. The best datasets were collected from a crystal soaked for 1 hour. The strategies used are listed in **Table V.I**.

Table V.I: Strategies for dataset collection for the best diffracting crystal.
Hrm refers to the high remote dataset.

Dataset	Beam-line	Energy (eV)	Omega start	Rotation	Number of images	Exposure (seconds)	Resolution (Å)	Trans-mission (%)
Peak	I02	9880.9	45˚	0.15˚	950	0.1	3.5	15
Inflection	I02	9877.4	45˚	0.15˚	950	0.1	3.5	15
Hrm	I02	9890.0	45˚	0.15˚	950	0.1	3.5	15

The data was then processed in XDS as described earlier for the anomalous crystal and the three datasets were scaled together using XSCALE (Kabsch, 2010). The heavy atom sites were identified with SHELXC/D/E (Sheldrick, 2010) and entered into autoSHARP (Vonrhein et al, 2007) where the phasing was performed according to manufacturer's instructions using default settings for multi-wavelength anomalous dispersion (MAD) data.

Results

The parameters obtained for each dataset after processing can be seen in **Table V.II**. A fluorescence scan was performed on the soaked crystals and can be seen in **Figure V.I**. The resulting electron density map is shown in the main text in **Figure 6.10**.

Table V.II: Summary of the parameters obtained after the initial processing of the three datasets obtained suing the tantalum bromide soaked crystals.

The numbers in brackets correspond to the values for the highest resolution shell. The anomalous signal represents the mean difference between the structure factor estimates in numbers of standard deviations.

Dataset	Peak	Inflection	High Remote
Beamline	I02	I02	I02
Energy (eV)	9880.9	9877.4	9890.0
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit Cell	a = 65.04, b = 175.20, c = 319.18 $\alpha = \beta = \gamma = 90^\circ$	a = 65.78, b = 175.37, c = 318.01 $\alpha = \beta = \gamma = 90^\circ$	a = 65.54, b = 175.40, c = 318.00 $\alpha = \beta = \gamma = 90^\circ$
Resolution	48.62-4.20 (4.43-4.20) Å	48.62-4.20 (4.43-4.20) Å	48.62-4.20 (4.43-4.20) Å
Reflections	143,710	143,704	134,224
Unique Reflections	27,613	27,756	26,056
$R_{measured}$ (%)	18.5 (54.3)	18.9 (67.1)	18.1 (62.1)
Completeness (%)	99.0 (94.4)	99.3 (96.7)	99.2 (92.8)
I/σI	10.3 (3.2)	9.5 (2.7)	9.4 (2.6)
Multiplicity	5.2 (5.3)	5.2 (5.2)	5.2 (4.9)
Anomalous signal	1.46 (0.77)	1.15 (0.79)	1.06 (0.77)

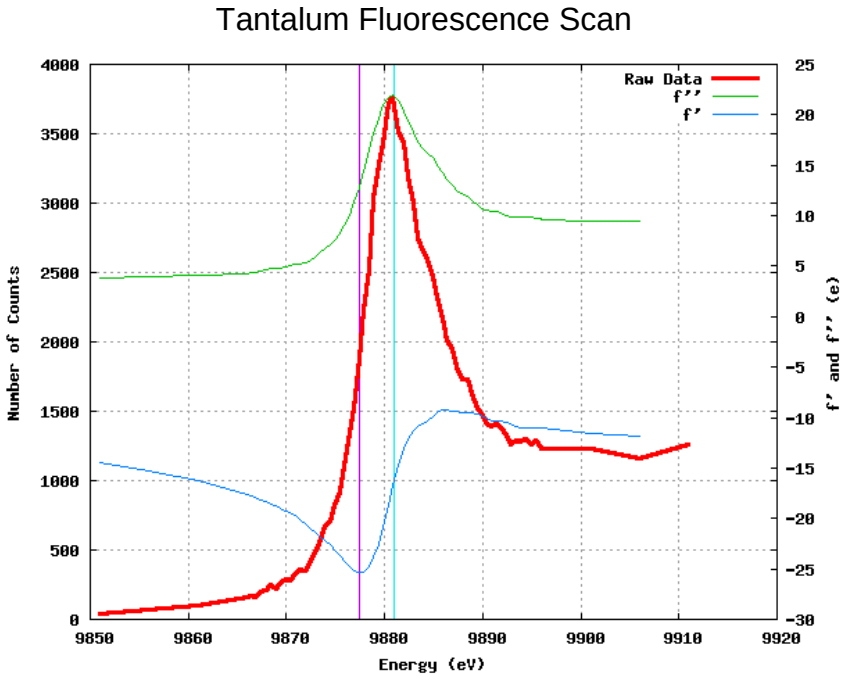


Figure V.I: Fluorescence scan obtained at the L3 edge of tantalum absorption using a soaked crystal.

The experimental data is shown as a red line (as shown before with the seleno-methionine protein). The peak (green) and inflection (blue) energies were calculated from the graph.

10 REFERENCES

Abergel C (2013) Molecular replacement: tricks and treats. *Acta Crystallogr D* **69:** 2167-2173

Amlacher S, Sarges P, Flemming D, van Noort V, Kunze R, Devos DP, Arumugam M, Bork P, Hurt E (2011) Insight into Structure and Assembly of the Nuclear Pore Complex by Utilizing the Genome of a Eukaryotic Thermophile. *Cell* **146:** 277-289

Arcus VL, Backbro K, Roos A, Daniel EL, Baker EN (2004) Distant structural homology leads to the functional characterization of an archaeal PIN domain as an exonuclease. *J Biol Chem* **279:** 16471-16478

Arts GJ, Englmeier L, Mattaj IW (1997) Energy-and temperature-dependent in vitro export of RNA from synthetic nuclei. *Biol Chem* **378:** 641-649

Aslanidis C, Dejong PJ (1990) Ligation-Independent Cloning of Pcr Products (Lic-Pcr). *Nucleic Acids Res* **18:** 6069-6074

Berne BJ, Pecora R (2000) *Dynamic light scattering : with applications to chemistry, biology, and physics*, Dover edn. Mineola, N.Y.: Dover Publications.

Blundell TL, Johnson LN (1976) *Protein Crystallogaphy*, London, UK: Academic Press Inc., Ltd.

Burchard W, Schmidt M, Stockmayer WH (1980) Information on Polydispersity and Branching from Combined Quasi-Elastic and Integrated Scattering. *Macromolecules* **13:** 1265-1272

Bursac S, Brdovcak MC, Donati G, Volarevic S (2014) Activation of the tumor suppressor p53 upon impairment of ribosome biogenesis. *Bba-Mol Basis Dis* **1842:** 817-830

Campbell MG, Karbstein K (2011) Protein-protein interactions within late pre-40S ribosomes. *Plos One* **6:** e16194

Carron C, O'Donohue MF, Choesmel V, Faubladier M, Gleizes PE (2011) Analysis of two human pre-ribosomal factors, bystin and hTsr1, highlights differences in evolution of ribosome biogenesis between yeast and mammals. *Nucleic Acids Res* **39:** 280-291

Chen W, Bucaria J, Band DA, Sutton A, Sternglanz R (2003) Enp1, a yeast protein associated with U3 and U14 snoRNAs, is required for pre-rRNA processing and 40S subunit synthesis. *Nucleic Acids Res* **31:** 690-699

Connolly K, Culver G (2009) Deconstructing ribosome construction. *Trends Biochem Sci* **34:** 256-263

Datta PP, Wilson DN, Kawazoe M, Swami NK, Kaminishi T, Sharma MR, Booth TM, Takemoto C, Fucini P, Yokoyama S, Agrawal RK (2007) Structural aspects of RbfA action during small ribosomal subunit assembly. *Mol Cell* **28:** 434-445

Derewenda ZS (2004) The use of recombinant methods and molecular engineering in protein crystallization. *Methods* **34:** 354-363

Dragon F, Gallagher JE, Compagnone-Post PA, Mitchell BM, Porwancher KA, Wehner KA, Wormsley S, Settlage RE, Shabanowitz J, Osheim Y, Beyer AL, Hunt DF, Baserga SJ (2002) A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. *Nature* **417:** 967-970

Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta crystallographica Section D, Biological crystallography* **66:** 486-501

Fassio CA, Schofield BJ, Seiser RM, Johnson AW, Lycan DE (2010) Dominant mutations in the late 40S biogenesis factor Ltv1 affect cytoplasmic maturation of the small ribosomal subunit in Saccharomyces cerevisiae. *Genetics* **185:** 199-209

Fatica A, Oeffinger M, Dlakic M, Tollervey D (2003) Nob1p is required for cleavage of the 3' end of 18S rRNA. *Mol Cell Biol* **23:** 1798-1807

Fatica A, Tollervey D, Dlakic M (2004) PIN domain of Nob1p is required for D-site cleavage in 20S pre-rRNA. *RNA* **10:** 1698-1701

Feig AL (2007) Applications of isothermal titration calorimetry in RNA biochemistry and biophysics. *Biopolymers* **87:** 293-301

Ferreira-Cerca S, Sagar V, Schafer T, Diop M, Wesseling AM, Lu H, Chai E, Hurt E, LaRonde-LeBlanc N (2012) ATPase-dependent role of the atypical kinase Rio2 on the evolving pre-40S ribosomal subunit. *Nature structural & molecular biology* **19:** 1316-1323

Freed EF, Bleichert F, Dutca LM, Baserga SJ (2010) When ribosomes go bad: diseases of ribosome biogenesis. *Mol Biosyst* **6:** 481-493

Gelperin D, Horton L, Beckman J, Hensold J, Lemmon SK (2001) Bms1p, a novel GTP-binding protein, and the related Tsr1p are required for distinct steps of 40S ribosome biogenesis in yeast. *RNA* **7:** 1268-1283

Golomb L, Volarevic S, Oren M (2014) p53 and ribosome biogenesis stress: The essentials. *Febs Letters* **588:** 2571-2579

Granneman S, Nandineni MR, Baserga SJ (2005) The putative NTPase Fap7 mediates cytoplasmic 20S pre-rRNA processing through a direct interaction with Rps14. *Mol Cell Biol* **25:** 10352-10364

Granneman S, Petfalski E, Swiatkowska A, Tollervey D (2010) Cracking pre-40S ribosomal subunit structure by systematic analyses of RNA-protein cross-linking. *EMBO J* **29:** 2026-2036

Green R, Noller HF (1997) Ribosomes and translation. *Annual review of biochemistry* **66:** 679-716

Grimm C, Chari A, Reuter K, Fischer U (2010) A crystallization screen based on alternative polymeric precipitants. *Acta crystallographica Section D, Biological crystallography* **66:** 685-697

Hage AE, Tollervey D (2004) A surfeit of factors: why is ribosome assembly so much more complicated in eukaryotes than bacteria? *RNA Biol* **1:** 10-15

Hahn TE (1995) *International Tables for Crystallography, Volume A*, Dordrecht, Netherlands: Kluwer Academic Publishers.

Hellmich UA, Weis BL, Lioutikov A, Wurm JP, Kaiser M, Christ NA, Hantke K, Kotter P, Entian KD, Schleiff E, Wohnert J (2013) Essential ribosome assembly factor Fap7 regulates a hierarchy of RNA-protein interactions during small ribosomal subunit biogenesis. *Proc Natl Acad Sci U S A* **110:** 15253-15258

Henras AK, Soudet J, Gerus M, Lebaron S, Caizergues-Ferrer M, Mougin A, Henry Y (2008) The post-transcriptional steps of eukaryotic ribosome biogenesis. *Cell Mol Life Sci* **65:** 2334-2359

Incardona MF, Bourenkov GP, Levik K, Pieritz RA, Popov AN, Svensson O (2009) EDNA: a framework for plugin-based applications applied to X-ray experiment online data analysis. *Journal of synchrotron radiation* **16:** 872-879

Inoue K, Alsina J, Chen JQ, Inouye M (2003) Suppression of defective ribosome assembly in a rbfA deletion mutant by overexpression of Era, an essential GTPase in Escherichia coli. *Mol Microbiol* **48:** 1005-1016

Jancarik J, Kim SH (1991) Sparse-Matrix Sampling - a Screening Method for Crystallization of Proteins. *Journal of applied crystallography* **24:** 409-411

Jensen BC, Brekken DL, Randall AC, Kifer CT, Parsons M (2005) Species specificity in ribosome biogenesis: a nonconserved phosphoprotein is required for formation of the large ribosomal subunit in Trypanosoma brucei. *Eukaryot Cell* **4:** 30-35

Jia MZ, Horita S, Nagata K, Tanokura M (2010) An archaeal Dim2-like protein, aDim2p, forms a ternary complex with a/eIF2 alpha and the 3' end fragment of 16S rRNA. *J Mol Biol* **398:** 774-785

Jia MZ, Ohtsuka J, Lee WC, Nagata K, Tanokura M (2007) Crystal structure of Dim2p: a preribosomal RNA processing factor, from Pyrococcus horikoshii OT3 at 2.30 A. *Proteins* **69:** 428-432

Kabsch W (2010) Xds. *Acta crystallographica Section D, Biological crystallography* **66:** 125-132

Kantardjieff KA, Rupp B (2003a) Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci* **12:** 1865-1871

Kantardjieff KA, Rupp B (2003b) Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci* **12:** 1865-1871

Karathia H, Vilaprinyo E, Sorribas A, Alves R (2011) Saccharomyces cerevisiae as a Model Organism: A Comparative Study. *Plos One* **6:** e16015

Kolakofsky D, Ohta T, Thach RE (1968) Junction of the 50S ribosomal subunit with the 30S initiation complex. *Nature* **220:** 244-247

Kos M, Tollervey D (2010) Yeast pre-rRNA processing and modification occur cotranscriptionally. *Mol Cell* **37:** 809-820

Kressler D, Hurt E, Bassler J (2010) Driving ribosome assembly. *Biochim Biophys Acta* **1803:** 673-683

Krezel A, Lesniak W, Jezowska-Bojczuk M, Mlynarz P, Brasun J, Kozlowski H, Bal W (2001) Coordination of heavy metals by dithiothreitol, a commonly used thiol group protectant. *J Inorg Biochem* **84:** 77-88

Kuhle B, Ficner R (2014) eIF5B employs a novel domain release mechanism to catalyze ribosomal subunit joining. *EMBO J* **33:** 1177-1191

Lafontaine D, Delcour J, Glasser AL, Desgres J, Vandenhaute J (1994) The DIM1 gene responsible for the conserved m6(2)Am6(2)A dimethylation in the 3'-terminal loop of 18 S rRNA is essential in yeast. *J Mol Biol* **241:** 492-497

Lamanna AC, Karbstein K (2009) Nob1 binds the single-stranded cleavage site D at the 3'-end of 18S rRNA with its PIN domain. *Proc Natl Acad Sci U S A* **106:** 14259-14264

Lamanna AC, Karbstein K (2011) An RNA conformational switch regulates pre-18S rRNA cleavage. *J Mol Biol* **405:** 3-17

LaRonde-LeBlanc N, Wlodawer A (2004) Crystal structure of A. fulgidus Rio2 defines a new family of serine protein kinases. *Structure* **12:** 1585-1594

Lebaron S, Papin C, Capeyrou R, Chen YL, Froment C, Monsarrat B, Caizergues-Ferrer M, Grigoriev M, Henry Y (2009) The ATPase and helicase activities of Prp43p are stimulated by the G-patch protein Pfa1p during yeast ribosome biogenesis. *EMBO J* **28:** 3808-3819

Lebaron S, Schneider C, van Nues RW, Swiatkowska A, Walsh D, Bottcher B, Granneman S, Watkins NJ, Tollervey D (2012) Proofreading of pre-40S ribosome maturation by a translation initiation factor and 60S subunits. *Nature structural & molecular biology* **19:** 744-753

Leibundgut M, Frick C, Thanbichler M, Bock A, Ban N (2005) Selenocysteine tRNA-specific elongation factor SelB is a structural chimaera of elongation and initiation factors. *EMBO J* **24:** 11-22

Li Y, Ma C, Qian M, Wen Z, Jing H, Qian D (2014) Downregulation of NOB1 suppresses the proliferation and tumor growth of non-small cell lung cancer in vitro and in vivo. *Oncology reports* **31:** 1271-1276

Lin Y, Peng S, Yu H, Teng H, Cui M (2011) RNAi-mediated downregulation of NOB1 suppresses the growth and colony-formation ability of human ovarian cancer cells. *Med Oncol* **29:** 311-317

Loc'h J, Blaud M, Rety S, Lebaron S, Deschamps P, Bareille J, Jombart J, Robert-Paganin J, Delbos L, Chardon F, Zhang E, Charenton C, Tollervey D, Leulliot N (2014) RNA mimicry by the fap7 adenylate kinase in ribosome biogenesis. *PLoS Biol* **12:** e1001860

Lu Z, Guo Q, Shi A, Xie F, Lu Q (2011) Downregulation of NIN/RPN12 binding protein inhibit the growth of human hepatocellular carcinoma cells. *Mol Biol Rep* **39:** 501-507

Matthews BW (1968) Solvent Content of Protein Crystals. *J Mol Biol* **33:** 491-&

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *Journal of applied crystallography* **40:** 658-674

McLellan T (1982) Electrophoresis buffers for polyacrylamide gels at various pH. *Analytical biochemistry* **126:** 94-99

Narla A, Ebert BL (2010) Ribosomopathies: human disorders of ribosome dysfunction. *Blood* **115:** 3196-3205

Newman J, Egan D, Walter TS, Meged R, Berry I, Ben Jelloul M, Sussman JL, Stuart DI, Perrakis A (2005) Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the PACT/JCSG+ strategy. *Acta crystallographica Section D, Biological crystallography* **61:** 1426-1431

Nissen P, Hansen J, Ban N, Moore PB, Steitz TA (2000) The structural basis of ribosome activity in peptide bond synthesis. *Science* **289:** 920-930

Noller HF, Hoffarth V, Zimniak L (1992) Unusual resistance of peptidyl transferase to protein extraction procedures. *Science* **256:** 1416-1419

O'Farrell HC, Pulicherla N, Desai PM, Rife JP (2006) Recognition of a complex substrate by the KsgA/Dim1 family of enzymes has been conserved throughout evolution. *RNA* **12:** 725-733

O'Farrell HC, Scarsdale JN, Rife JP (2004) Crystal structure of KsgA, a universally conserved rRNA adenine dimethyltransferase in Escherichia coli. *J Mol Biol* **339:** 337-353

Oliva A, Llabres M, Farina JB (2001) Comparative study of protein molecular weights by size-exclusion chromatography and laser-light scattering. *Journal of pharmaceutical and biomedical analysis* **25:** 833-841

Ortega A, Amoros D, de la Torre JG (2011) Prediction of Hydrodynamic and Other Solution Properties of Rigid Proteins from Atomic- and Residue-Level Models. *Biophys J* **101:** 892-898

Osheim YN, French SL, Keck KM, Champion EA, Spasov K, Dragon F, Baserga SJ, Beyer AL (2004) Pre-18S ribosomal RNA is structurally compacted into the SSU processome prior to being cleaved from nascent transcripts in Saccharomyces cerevisiae. *Mol Cell* **16:** 943-954

Pertschy B, Schneider C, Gnadig M, Schafer T, Tollervey D, Hurt E (2009) RNA helicase Prp43 and its co-factor Pfa1 promote 20 to 18 S rRNA processing catalyzed by the endonuclease Nob1. *J Biol Chem* **284:** 35079-35091

Prohaska K, Williams N (2009) Assembly of the Trypanosoma brucei 60S ribosomal subunit nuclear export complex requires trypanosome-specific proteins P34 and P37. *Eukaryot Cell* **8:** 77-87

Pulicherla N, Pogorzala LA, Xu Z, HC OF, Musayev FN, Scarsdale JN, Sia EA, Culver GM, Rife JP (2009) Structural and functional divergence within the Dim1/KsgA family of rRNA methyltransferases. *J Mol Biol* **391:** 884-893

Radaev S, Li S, Sun PD (2006) A survey of protein-protein complex crystallizations. *Acta crystallographica Section D, Biological crystallography* **62:** 605-612

Rouquette J, Choesmel V, Gleizes PE (2005) Nuclear export and cytoplasmic processing of precursors to the 40S ribosomal subunits in mammalian cells. *EMBO J* **24:** 2862-2872

Schafer T, Maco B, Petfalski E, Tollervey D, Bottcher B, Aebi U, Hurt E (2006) Hrr25-dependent phosphorylation state regulates organization of the pre-40S subunit. *Nature* **441:** 651-655

Schafer T, Strauss D, Petfalski E, Tollervey D, Hurt E (2003) The path from nucleolar 90S to cytoplasmic 40S pre-ribosomes. *EMBO J* **22:** 1370-1380

Seiser RM, Sundberg AE, Wollam BJ, Zobel-Thropp P, Baldwin K, Spector MD, Lycan DE (2006) Ltv1 is required for efficient nuclear export of the ribosomal small subunit in Saccharomyces cerevisiae. *Genetics* **174:** 679-691

Sheldrick GM (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta crystallographica Section D, Biological crystallography* **66:** 479-485

Stein N (2008) CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *Journal of applied crystallography* **41:** 641-643

Storoni LC, McCoy AJ, Read RJ (2004) Likelihood-enhanced fast rotation functions. *Acta crystallographica Section D, Biological crystallography* **60:** 432-438

Strunk BS, Loucks CR, Su M, Vashisth H, Cheng S, Schilling J, Brooks CL, 3rd, Karbstein K, Skiniotis G (2011) Ribosome assembly factors prevent premature translation initiation by 40S assembly intermediates. *Science* **333:** 1449-1453

Strunk BS, Novak MN, Young CL, Karbstein K (2012) A translation-like cycle is a quality control checkpoint for maturing 40S ribosome subunits. *Cell* **150:** 111-121

Tone Y, Toh EA (2002) Nob1p is required for biogenesis of the 26S proteasome and degraded upon its maturation in Saccharomyces cerevisiae. *Genes Dev* **16:** 3142-3157

Traub P, Nomura M (1968) Structure and function of E. coli ribosomes. V. Reconstitution of functionally active 30S ribosomal particles from RNA and proteins. *Proc Natl Acad Sci U S A* **59:** 777-784

Turowski TW, Tollervey D (2014) Cotranscriptional events in eukaryotic ribosome synthesis. *Wiley interdisciplinary reviews RNA*

Vandesande W, Persoons A (1985) The Size and Shape of Macromolecular Structures - Determination of the Radius, the Length, and the Persistence Length of Rodlike Micelles of Dodecyldimethylammonium Chloride and Bromide. *J Phys Chem-Us* **89:** 404-406

Vanrobays E, Gelugne JP, Caizergues-Ferrer M, Lafontaine DL (2004) Dim2p, a KH-domain protein required for small ribosomal subunit synthesis. *RNA* **10:** 645-656

Vanrobays E, Gelugne JP, Gleizes PE, Caizergues-Ferrer M (2003) Late cytoplasmic maturation of the small ribosomal subunit requires RIO proteins in Saccharomyces cerevisiae. *Mol Cell Biol* **23:** 2083-2095

Vanrobays E, Leplus A, Osheim YN, Beyer AL, Wacheul L, Lafontaine DL (2008) TOR regulates the subcellular distribution of DIM2, a KH domain protein required for cotranscriptional ribosome assembly and pre-40S ribosome export. *RNA* **14:** 2061-2073

Veith T, Martin R, Wurm JP, Weis BL, Duchardt-Ferner E, Safferthal C, Hennig R, Mirus O, Bohnsack MT, Wohnert J, Schleiff E (2012) Structural and functional analysis of the archaeal endonuclease Nob1. *Nucleic Acids Res* **40:** 3259-3274

Vonrhein C, Blanc E, Roversi P, Bricogne G (2007) Automated structure solution with autoSHARP. *Methods Mol Biol* **364:** 215-230

Wang X, Wu T, Hu Y, Marcinkiewicz M, Qi S, Valderrama-Carvajal H, Luo H, Wu J (2012) Pno1 tissue-specific expression and its functions related to the immune responses and proteasome activities. *Plos One* **7:** e46093

Wen J, Arakawa T, Philo JS (1996) Size-exclusion chromatography with on-line light-scattering, absorbance, and refractive index detectors for studying proteins and their interactions. *Analytical biochemistry* **240:** 155-166

Widmann B, Wandrey F, Badertscher L, Wyler E, Pfannstiel J, Zemp I, Kutay U (2012) The kinase activity of human Rio1 is required for final steps of cytoplasmic maturation of 40S subunits. *Molecular biology of the cell* **23:** 22-35

Wild T, Horvath P, Wyler E, Widmann B, Badertscher L, Zemp I, Kozak K, Csucs G, Lund E, Kutay U (2010) A protein inventory of human ribosome biogenesis reveals an essential function of exportin 5 in 60S subunit export. *PLoS Biol* **8:** e1000522

Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS (2011) Overview of the CCP4 suite and current developments. *Acta crystallographica Section D, Biological crystallography* **67:** 235-242

Woolls HA, Lamanna AC, Karbstein K (2011) Roles of Dim2 in ribosome assembly. *J Biol Chem* **286:** 2578-2586

Zemp I, Wild T, O'Donohue MF, Wandrey F, Widmann B, Gleizes PE, Kutay U (2009) Distinct cytoplasmic maturation steps of 40S ribosomal subunit precursors require hRio2. *J Cell Biol* **185:** 1167-1180

Zhou J, Xu T, Yan Y, Qin R, Wang H, Zhang X, Huang Y, Wang Y, Lu Y, Fu D, Chen J (2013) MicroRNA-326 functions as a tumor suppressor in glioma by targeting the Nin one binding protein (NOB1). *Plos One* **8:** e68469